

MAIZE DEFENSE-INDUCIBLE GENES AND THEIR USE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of, and claims benefit of and priority to U.S. Application Serial No. 10/027,559, filed on October 23, 2001, which claims benefit of and
5 priority to U.S. Provisional Patent Application No. 60/243,120, filed on October 25, 2000, both of which are hereby incorporated by reference in their entirety.

TECHNICAL FIELD

The present invention relates generally to plant molecular biology. More specifically,
10 it relates to nucleic acids and methods for modulating their expression in plants and to transforming genes into plants in order to enhance disease resistance.

BACKGROUND OF THE INVENTION

Disease in plants is caused by biotic and abiotic causes. Biotic causes include fungi,
15 viruses, insects, bacteria, and nematodes. Of these, fungi are the most frequent causative agents of disease in plants. Abiotic causes of disease in plants include extremes of temperature, water, oxygen, soil pH, plus nutrient-element deficiencies and imbalances, excess heavy metals, and air pollution.

A host of cellular processes enables plants to defend themselves from disease caused
20 by pathogenic agents. These processes apparently form an integrated set of resistance mechanisms that is activated by initial infection and then limits further spread of the invading pathogenic microorganism.

Subsequent to recognition of a potentially pathogenic microbe, plants can activate an array of biochemical responses. Generally, the plant responds by inducing several local
25 responses in the cells immediately surrounding the infection site. The most common resistance response observed in both nonhost and race-specific interactions is termed the

"hypersensitive response" (HR). In the hypersensitive response, cells contacted by the pathogen, and often neighboring cells, rapidly collapse and dry in a necrotic fleck. Other responses include the deposition of callose, the physical thickening of cell walls by lignification, and the synthesis of various antibiotic small molecules and proteins. Genetic factors in both the host and the pathogen determine the specificity of these local responses, which can be very effective in limiting the spread of infection.

As noted, among the causative agents of infectious disease of crop plants, the phytopathogenic fungi play the dominant role. Phytopathogenic fungi cause devastating epidemics, as well as causing significant annual crop yield losses. Pathogenic fungi attack all of the approximately 300,000 species of flowering plants. However, a single plant species can be host to only a few fungal species, and similarly, most fungi usually have a limited host range.

Plant disease outbreaks have resulted in catastrophic crop failures that have triggered famines and caused major social change. Generally, the best strategy for plant disease control is to use resistant cultivars selected or developed by plant breeders for this purpose. However, the potential for serious crop disease epidemics persists today, as evidenced by outbreaks of the Victoria blight of oats and southern corn leaf blight. Accordingly, molecular methods are needed to supplement traditional breeding methods to protect plants from pathogen attack.

SUMMARY OF THE INVENTION

Nucleic acids and proteins relating to defense-inducible genes in plants are provided. In particular, six defense-inducible nucleic acid and protein sequences are provided. The nucleic acid sequences can be used to alter the level, tissue, or timing of expression of the plant genes to achieve enhanced disease resistance. Transgenic plants comprising the nucleic acids of the present invention are also provided. Methods for modulating the expression of the nucleic acids in a transgenic plant are additionally disclosed.

Therefore, in one aspect, the present invention relates to an isolated nucleic acid comprising a member selected from the group consisting of (a) a polynucleotide encoding a polypeptide of the present invention; (b) a polynucleotide amplified from a *Zea mays* nucleic acid library using the primers of the present invention; (c) a polynucleotide comprising at least 20 contiguous bases of the polynucleotides of the present invention; (d) a polynucleotide encoding a plant defense-inducible protein; (e) a polynucleotide having at least 50% sequence

identity to the polynucleotides of the present invention; (f) a polynucleotide comprising at least 25 nucleotide in length which hybridizes under low stringency conditions to the polynucleotides of the present invention; and (g) a polynucleotide complementary to a polynucleotide of (a) through (f). The isolated nucleic acid can be DNA. The isolated nucleic acid can also be RNA.

In another aspect, the present invention relates to vectors comprising the polynucleotides of the present invention. Also, the present invention relates to recombinant expression cassettes, comprising a nucleic acid of the present invention operably linked to a promoter. In addition, the present invention relates to recombinant expression cassettes.

In another aspect, the present invention is directed to a host cell into which has been introduced the recombinant expression cassette.

In yet another aspect, the present invention relates to a transgenic plant or plant cell comprising a recombinant expression cassette with a promoter operably linked to any of the isolated nucleic acids of the present invention. Preferred plants containing the recombinant expression cassette of the present invention include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice barley, and millet. The present invention also provides transgenic seed from the transgenic plant.

In another aspect, the present invention relates to an isolated protein selected from the group consisting of (a) a polypeptide comprising at least 25 contiguous amino acids of SEQ ID NOS:2, 4, 6, 8, 10, and 12; (b) a polypeptide which is a plant defense-inducible protein; (c) a polypeptide comprising at least 55% sequence identity to SEQ ID NOS:2, 4, 6, 8, 10, and 12; (d) a polypeptide encoded by a nucleic acid of the present invention; (e) a polypeptide characterized by SEQ ID NOS:2, 4, 6, 8, 10, and 12; and (f) a conservatively modified variant of SEQ ID NOS:2, 4, 6, 8, 10, and 12.

In a further aspect, the present invention relates to a method of modulating the level of protein in a plant by introducing into a plant cell a recombinant expression cassette comprising a polynucleotide of the present invention operably linked to a promoter; culturing the plant cell under plant growing conditions to produce a regenerated plant; and inducing expression of the polynucleotide for a time sufficient to modulate the protein of the present invention in the plant. Preferred plants of the present invention include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. The level of protein in the plant can either be increased or decreased.

DETAILED DESCRIPTION OF THE INVENTION

Plant defense-inducible genes and polypeptides are provided. In particular, six defense-inducible sequences from maize are provided. By "defense-inducible" is intended
5 that the expression of the gene is increased or induced when the plant is responding to biotic and abiotic stress such as pathogen attack. That is, there may be increased mRNA production for the genes, greater corresponding protein product levels, as well as greater activity of the protein product. The nucleic acid sequences of the invention find use in conferring enhanced resistance to a plant. Thus, the sequences may be used to increase the resistance or tolerance
10 to known crop plant pathogens, including fungi, bacteria, viruses, other microbes, nematodes, insects and the like. Additionally, the sequences may confer resistance or tolerance to diseases caused by heat, drought, cold, reactive oxygen species, and radiation.

The present invention provides, among other things, compositions and methods for modulating (i.e., increasing or decreasing) the level of polynucleotides and polypeptides of the
15 present invention in plants. In particular, the polynucleotides and polypeptides of the present invention can be expressed temporally or spatially, e.g., at developmental stages, in tissues, and/or in quantities, which are uncharacteristic of non-recombinantly engineered plants. Thus, the present invention provides utility in such exemplary applications as enhanced disease resistance in plants.

20 In particular, six sequences are provided. The sequences were identified based on a blast search for related sequences in the public database. The sequences are selected from an extensin-like sequence (SEQ ID NOS:1 and 2), a cytosolic ascorbate peroxidase-like sequence (SEQ ID NOS:5 and 6), a metallothionein-like sequence (SEQ ID NOS:3 and 4), a peroxidase-like sequence (SEQ ID NOS:11 and 12), a non-specific lipid transfer protein-like
25 sequence (SEQ ID NOS:7 and 8), and a proteinase inhibitor-like sequence (SEQ ID NOS:9 and 10).

Extensin-like sequences are characterized by encoding a putative hydroxyproline-rich glycoproteins. The polypeptides generally have a high proportion of Pro, Lys, and Thr residues. The genes function in controlling the integrity, strength, and impenetrability of the
30 cell wall. As such, its increased expression is likely adaptive in that it improves the plants ability to ward off successful infection by plant pathogens by the increased integrity of the cell wall.

In plants, ascorbate peroxidase (APX) is an important peroxide-detoxifying enzyme. The expression of APX is rapidly induced in response to stresses that result in the accumulation of reactive oxygen species. The steady-state level of transcripts encoding cytosolic APX is dramatically induced during the hypersensitive response of plants infected with virus. Tolerance to low temperature and oxidative stress has been demonstrated for plants having increased ascorbate peroxidase activities. In general, ascorbate peroxidase has been suggested as a particularly important antioxidant enzyme in helping plants survive oxidative stress.

Plant metallothionein (MT) it is proposed sequesters excess copper, and possibly zinc preventing adverse metal-protein interactions. At least two different MT-like proteins have been identified in plants. MT-1 displays a Cys-X-Cys motif for all Cys residues, while MT-2 has the typical structure having Cys-Cys and Cys-X-X-Cys motifs within the N-terminal domain. The MT proteins are typically regulated by the developmental stage and may participate in the cell maturation process. The MT-like genes of the invention is predicted to encode a metal binding polypeptide. As these genes are defense-inducible, they may serve a function in conditioning the plant cells to be more resistant to pathogens, for example, by robbing pathogens of necessary metal cations that they use to live, grow, and gain access to the plant to cause disease. Thus, increasing metallothionein expression increases resistance to the pathogen in the plant.

The induction of defense-related peroxidase (POD) activity in plants occurs in response to many biotic and abiotic stimuli. In one study, exposure of seedlings to daily periods of wind induced a significant and sustained increase in soluble POD activity in primary leaves of seedlings. Thus, wind and other mechanical stimuli can act as inducers of POD activity and interacting factors in the elicitation of POD activity by other environmental stimuli. Induction in POD activity has also been observed in response to bacterization in plants. For example, POD activity was increased in roots following bacterization with *Pseudomonas*. The POD-like genes of the invention are involved in pathogen defense by controlling the level of reactive oxygen species in the cell.

Non-specific lipid transfer proteins show strong antifungal activity. The family of plant non-specific lipid transfer proteins share sequence homology including conserved cysteine residues. The proteins are expressed in plants prior to contact with a pathogen and are induced during infection and are present both intra- and extracellularly.

Immunohistological investigations have demonstrated that the proteins accumulate in contact with a fungal pathogen and are active in autolysing cells, suggesting a role in plant defense. The lipid transfer-like proteins of the invention have antimicrobial and antifungal function. The upregulation of the genes in defense situations suggests that the increased expression of an antipathogenic protein acts to increase resistance of the crop plant to pathogens.

Plant seeds contain a large number of protease inhibitors of animal, fungal, and bacterial origin. Other plant tissues also express protease inhibitors. Monocots have a 16K, double-headed inhibitor. The proteinase inhibitor-like proteins of the invention have antimicrobial and antifungal activity. The genes are induced during a defense response in plants. Thus, increased expression of the genes that encode the proteinase-like inhibitor proteins increase disease resistance in the plants.

The present invention also provides isolated nucleic acid comprising polynucleotides of sufficient length and complementarity to a gene of the present invention to use as probes or amplification primers in the detection, quantitation, or isolation of gene transcripts. For example, isolated nucleic acids of the present invention can be used as probes in detecting deficiencies in the level of mRNA in screenings for desired transgenic plants, for detecting mutations in the gene (e.g., substitutions, deletions, or additions), for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants (polymorphisms), orthologs, or paralogues of the gene, or for site directed mutagenesis in eukaryotic cells (see, e.g., U.S. Patent No. 5,565,350). The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides, or for use as immunogens in the preparation and/or screening of antibodies. The isolated nucleic acids of the present invention can also be employed for use in sense or antisense suppression of one or more genes of the present invention in a host cell, tissue, or plant. Attachment of chemical agents, which bind, intercalate, cleave and/or crosslink to the isolated nucleic acids of the present invention can also be used to modulate transcription or translation. The present invention also provides isolated proteins comprising a polypeptide of the present invention (e.g., preproenzyme, proenzyme, or enzymes).

The isolated nucleic acids and proteins of the present invention can be used over a broad range of plant types, particularly monocots such as the species of the family *Gramineae* including *Sorghum* (e.g. *S. bicolor*), *Oryza*, *Avena*, *Hordeum*, *Secale*, *Triticum* and *Zea mays*, and dicots such as *Glycine*. The isolated nucleic acid and proteins of the present invention can

also be used in species from the genera: *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*,
5 *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Pisum*, *Phaseolus*, *Lolium*, and *Allium*.

The invention is drawn to compositions and methods for inducing resistance in a plant to plant pests. Accordingly, the compositions and methods are also useful in protecting plants
10 against fungal pathogens, viruses, nematodes, insects and the like.

By "disease resistance" is intended that the plants avoid the disease symptoms that are the outcome of plant-pathogen interactions. That is, pathogens are prevented from causing plant diseases and the associated disease symptoms, or alternatively, the disease symptoms caused by the pathogen is minimized or lessened. Consequently, the sequence of the
15 invention find use in modulating (i.e., increasing or decreasing) disease resistance in a plant.

By "antipathogenic compositions" is intended that the compositions of the invention have antipathogenic activity and thus are capable of suppressing, controlling, and/or killing the invading pathogenic organism. An antipathogenic composition of the invention will reduce the disease symptoms resulting from pathogen challenge by at least about 5% to about
20 50%, at least about 10% to about 60%, at least about 30% to about 70%, at least about 40% to about 80%, or at least about 50% to about 90% or greater. Hence, the methods of the invention can be utilized to protect plants from disease, particularly those diseases that are caused by plant pathogens.

Assays that measure antipathogenic activity are commonly known in the art, as are
25 methods to quantitate disease resistance in plants following pathogen infection. See, for example, U.S. Patent No. 5,614,395, herein incorporated by reference. Such techniques include, measuring over time, the average lesion diameter, the pathogen biomass, and the overall percentage of decayed plant tissues. For example, a plant either expressing an antipathogenic polypeptide or having an antipathogenic composition applied to its surface
30 shows a decrease in tissue necrosis (i.e., lesion diameter) or a decrease in plant death following pathogen challenge when compared to a control plant that was not exposed to the antipathogenic composition. Alternatively, antipathogenic activity can be measured by a

decrease in pathogen biomass. For example, a plant expressing an antipathogenic polypeptide or exposed to an antipathogenic composition is challenged with a pathogen of interest. Over time, tissue samples from the pathogen-inoculated tissues are obtained and RNA is extracted. The percent of a specific pathogen RNA transcript relative to the level of a plant specific transcript allows the level of pathogen biomass to be determined. See, for example, Thomma
5 *et al.* (1998) *Plant Biology* 95:15107-15111, herein incorporated by reference.

Furthermore, *in vitro* antipathogenic assays include, for example, the addition of varying concentrations of the antipathogenic composition to paper disks and placing the disks on agar containing a suspension of the pathogen of interest. Following incubation, clear
10 inhibition zones develop around the discs that contain an effective concentration of the antipathogenic polypeptide (Liu *et al.* (1994) *Plant Biology* 91:1888-1892, herein incorporated by reference). Additionally, microspectrophotometrical analysis can be used to measure the *in vitro* antipathogenic properties of a composition (Hu *et al.* (1997) *Plant Mol. Biol.* 34:949-959 and Cammue *et al.* (1992) *J. Biol. Chem.* 267:2228-2233, both of which are
15 herein incorporated by reference).

Pathogens of the invention include, but are not limited to, viruses or viroids, bacteria, insects, nematodes, fungi, and the like. Viruses include any plant virus, for example, tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Specific fungal and viral pathogens for the major crops include: Soybeans: *Phytophthora*
20 *megasperma* fsp. *glycinea*, *Macrophomina phaseolina*, *Rhizoctonia solani*, *Sclerotinia sclerotiorum*, *Fusarium oxysporum*, *Diaporthe phaseolorum* var. *sojae* (*Phomopsis sojae*), *Diaporthe phaseolorum* var. *caulivora*, *Sclerotium rolfsii*, *Cercospora kikuchii*, *Cercospora sojae*, *Peronospora manshurica*, *Colletotrichum dematium* (*Colletotrichum truncatum*), *Corynespora cassiicola*, *Septoria glycines*, *Phyllosticta sojicola*, *Alternaria alternata*,
25 *Pseudomonas syringae* p.v. *glycinea*, *Xanthomonas campestris* p.v. *phaseoli*, *Microsphaera diffusa*, *Fusarium semitectum*, *Phialophora gregata*, Soybean mosaic virus, *Glomerella glycines*, Tobacco Ring spot virus, Tobacco Streak virus, *Phakopsora pachyrhizi*, *Pythium aphanidermatum*, *Pythium ultimum*, *Pythium debaryanum*, Tomato spotted wilt virus, *Heterodera glycines* *Fusarium solani*; Canola: *Albugo candida*, *Alternaria brassicae*,
30 *Leptosphaeria maculans*, *Rhizoctonia solani*, *Sclerotinia sclerotiorum*, *Mycosphaerella brassicicola*, *Pythium ultimum*, *Peronospora parasitica*, *Fusarium roseum*, *Alternaria alternata*; Alfalfa: *Clavibacter michiganense* subsp. *insidiosum*, *Pythium ultimum*, *Pythium*

irregulare, *Pythium splendens*, *Pythium debaryanum*, *Pythium aphanidermatum*,
Phytophthora megasperma, *Peronospora trifoliorum*, *Phoma medicaginis* var. *medicaginis*,
Cercospora medicaginis, *Pseudopeziza medicaginis*, *Leptotrochila medicaginis*, *Fusarium-atrum*,
Xanthomonas campestris p.v. *alfalfae*, *Aphanomyces euteiches*, *Stemphylium herbarum*,
5 *Stemphylium alfalfae*; Wheat: *Pseudomonas syringae* p.v. *atrofaciens*, *Urocystis agropyri*,
Xanthomonas campestris p.v. *translucens*, *Pseudomonas syringae* p.v. *syringae*, *Alternaria*
alternata, *Cladosporium herbarum*, *Fusarium graminearum*, *Fusarium avenaceum*, *Fusarium*
culmorum, *Ustilago tritici*, *Ascochyta tritici*, *Cephalosporium gramineum*, *Collotetrichum*
graminicola, *Erysiphe graminis* f.sp. *tritici*, *Puccinia graminis* f.sp. *tritici*, *Puccinia recondita*
10 f.sp. *tritici*, *Puccinia striiformis*, *Pyrenophora tritici-repentis*, *Septoria nodorum*, *Septoria*
tritici, *Septoria avenae*, *Pseudocercospora herpotrichoides*, *Rhizoctonia solani*,
Rhizoctonia cerealis, *Gaeumannomyces graminis* var. *tritici*, *Pythium aphanidermatum*,
Pythium arrhenomanes, *Pythium ultimum*, *Bipolaris sorokiniana*, Barley Yellow Dwarf Virus,
Brome Mosaic Virus, Soil Borne Wheat Mosaic Virus, Wheat Streak Mosaic Virus, Wheat
15 Spindle Streak Virus, American Wheat Striate Virus, *Claviceps purpurea*, *Tilletia tritici*,
Tilletia laevis, *Ustilago tritici*, *Tilletia indica*, *Rhizoctonia solani*, *Pythium arrhenomanes*,
Pythium graminicola, *Pythium aphanidermatum*, High Plains Virus, European wheat striate
virus; Sunflower: *Plasmophora halstedii*, *Sclerotinia sclerotiorum*, Aster Yellows, *Septoria*
helianthi, *Phomopsis helianthi*, *Alternaria helianthi*, *Alternaria zinniae*, *Botrytis cinerea*,
20 *Phoma macdonaldii*, *Macrophomina phaseolina*, *Erysiphe cichoracearum*, *Rhizopus oryzae*,
Rhizopus arrhizus, *Rhizopus stolonifer*, *Puccinia helianthi*, *Verticillium dahliae*, *Erwinia*
carotovorum p.v. *Carotovora*, *Cephalosporium acremonium*, *Phytophthora cryptogea*, *Albugo*
tragopogonis; Maize: *Fusarium moniliforme* var. *subglutinans*, *Erwinia stewartii*, *Fusarium*
moniliforme, *Gibberella zeae* (*Fusarium graminearum*), *Stenocarpella maydis* (*Diplodia*
25 *maydis*), *Pythium irregulare*, *Pythium debaryanum*, *Pythium graminicola*, *Pythium splendens*,
Pythium ultimum, *Pythium aphanidermatum*, *Aspergillus flavus*, *Bipolaris maydis* O,T
(*Cochliobolus heterostrophus*), *Helminthosporium carbonum* I, II & III (*Cochliobolus*
carbonum), *Exserohilum turcicum* I, II & III, *Helminthosporium pedicellatum*, *Physoderma*
maydis, *Phyllosticta maydis*, *Kabatie-maydis*, *Cercospora sorghi*, *Ustilago maydis*, *Puccinia*
30 *sorghi*, *Puccinia polysora*, *Macrophomina phaseolina*, *Penicillium oxalicum*, *Nigrospora*
oryzae, *Cladosporium herbarum*, *Curvularia lunata*, *Curvularia inaequalis*, *Curvularia*
pallescens, *Clavibacter michiganense* subsp. *nebraskense*, *Trichoderma viride*, Maize Dwarf

- Mosaic Virus A & B, Wheat Streak Mosaic Virus, Maize Chlorotic Dwarf Virus, *Claviceps sorghi*, *Pseudomonas avenae*, *Erwinia chrysanthemi* p.v. *Zea*, *Erwinia corotovora*, *Cornstunt spiropasma*, *Diplodia macrospora*, *Sclerophthora macrospora*, *Peronosclerospora sorghi*, *Peronosclerospora philippinesis*, *Peronosclerospora maydis*, *Peronosclerospora sacchari*,
- 5 *Spacelotheca reiliana*, *Physopella zea*, *Cephalosporium maydis*, *Caphalosporium acremonium*, Maize Chlorotic Mottle Virus, High Plains Virus, Maize Mosaic Virus, Maize Rayado Fino Virus, Maize Streak Virus, Maize Stripe Virus, Maize Rough Dwarf Virus; Sorghum: *Exserohilum turcicum*, *Colletotrichum graminicola* (*Glomerella graminicola*), *Cercospora sorghi*, *Gloeocercospora sorghi*, *Ascochyta sorghina*, *Pseudomonas syringae* p.v.
- 10 *syringae*, *Xanthomonas campestris* p.v. *holcicola* *Pseudomonas andropogonis*, *Puccinia purpurea*, *Macrophomina phaseolina*, *Perconia circinata*, *Fusarium moniliforme*, *Alternaria alternate*, *Bipolaris sorghicola*, *Helminthosporium sorghicola*, *Curvularia lunata*, *Phoma insidiosa*, *Pseudomonas avenae* (*Pseudomonas alboprecipitans*), *Ramulispora sorghi*, *Ramulispora sorghicola*, *Phyllachara sacchari*, *Sporisorium reilianum* (*Sphacelotheca*
- 15 *reiliana*), *Sphacelotheca cruenta*, *Sporisorium sorghi*, Sugarcane mosaic H, Maize Dwarf Mosaic Virus A & B, *Claviceps sorghi*, *Rhizoctonia solani*, *Acremonium strictum*, *Sclerophthora macrospora*, *Peronosclerospora sorghi*, *Peronosclerospora philippinensis*, *Sclerospora graminicola*, *Fusarium graminearum*, *Fusarium oxysporum*, *Pythium arrhenomanes*, *Pythium graminicola*, etc.
- 20 Nematodes include parasitic nematodes such as root-knot, cyst, and lesion nematodes, including *Heterodera* and *Globodera* spp; particularly *Globodera rostochiensis* and *globodera pailida* (potato cyst nematodes); *Heterodera glycines* (soybean cyst nematode); *Heterodera schachtii* (beet cyst nematode); and *Heterodera avenae* (cereal cyst nematode).
- Insect pests include insects selected from the orders Coleoptera, Diptera,
- 25 Hymenoptera, Lepidoptera, Mallophaga, Homoptera, Hemiptera, Orthoptera, Thysanoptera, Dermaptera, Isoptera, Anoplura, Siphonaptera, Trichoptera, etc., particularly Coleoptera and Lepidoptera. Insect pests of the invention for the major crops include: Maize: *Ostrinia nubilalis*, European corn borer; *Agrotis ipsilon*, black cutworm; *Helicoverpa zea*, corn earworm; *Spodoptera frugiperda*, fall armyworm; *Diatraea grandiosella*, southwestern corn
- 30 borer; *Elasmopalpus lignosellus*, lesser cornstalk borer; *Diatraea saccharalis*, sugarcane borer; *Diabrotica virgifera*, western corn rootworm; *Diabrotica longicornis barberi*, northern corn rootworm; *Diabrotica undecimpunctata howardi*, southern corn rootworm; *Melanotus*

- spp.*, wireworms; *Cyclocephala borealis*, northern masked chafer (white grub); *Cyclocephala immaculata*, southern masked chafer (white grub); *Popillia japonica*, Japanese beetle; *Chaetocnema pulicaria*, corn flea beetle; *Sphenophorus maidis*, maize billbug; *Rhopalosiphum maidis*, corn leaf aphid; *Anuraphis maidiradicis*, corn root aphid; *Blissus leucopterus leucopterus*, chinch bug; *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus sanguinipes*, migratory grasshopper; *Hylemya platura*, seedcorn maggot; *Agromyza parvicornis*, corn blot leafminer; *Anaphothrips obscurus*, grass thrips; *Solenopsis milesta*, thief ant; *Tetranychus urticae*, twospotted spider mite; Sorghum: *Chilo partellus*, sorghum borer; *Spodoptera frugiperda*, fall armyworm; *Helicoverpa zea*, corn earworm;
- 10 *Elasmopalpus lignosellus*, lesser cornstalk borer; *Feltia subterranea*, granulate cutworm; *Phyllophaga crinita*, white grub; *Eleodes*, *Conoderus*, and *Aeolus spp.*, wireworms; *Oulema melanopus*, cereal leaf beetle; *Chaetocnema pulicaria*, corn flea beetle; *Sphenophorus maidis*, maize billbug; *Rhopalosiphum maidis*, corn leaf aphid; *Sipha flava*, yellow sugarcane aphid; *Blissus leucopterus leucopterus*, chinch bug; *Contarinia sorghicola*, sorghum midge;
- 15 *Tetranychus cinnabarinus*, carmine spider mite; *Tetranychus urticae*, twospotted spider mite; Wheat: *Pseudaletia unipunctata*, army worm; *Spodoptera frugiperda*, fall armyworm; *Elasmopalpus lignosellus*, lesser cornstalk borer; *Agrotis orthogonia*, western cutworm; *Elasmopalpus lignosellus*, lesser cornstalk borer; *Oulema melanopus*, cereal leaf beetle; *Hypera punctata*, clover leaf weevil; *Diabrotica undecimpunctata howardi*, southern corn
- 20 rootworm; Russian wheat aphid; *Schizaphis graminum*, greenbug; *Macrosiphum avenae*, English grain aphid; *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus differentialis*, differential grasshopper; *Melanoplus sanguinipes*, migratory grasshopper; *Mayetiola destructor*, Hessian fly; *Sitodiplosis mosellana*, wheat midge; *Meromyza americana*, wheat stem maggot; *Hylemya coarctata*, wheat bulb fly; *Frankliniella fusca*,
- 25 tobacco thrips; *Cephus cinctus*, wheat stem sawfly; *Aceria tulipae*, wheat curl mite; Sunflower: *Suleima helianthana*, sunflower bud moth; *Homoeosoma electellum*, sunflower moth; *zygogramma exclamationis*, sunflower beetle; *Bothyrus gibbosus*, carrot beetle; *Neolasioptera murtfeldtiana*, sunflower seed midge; Cotton: *Heliothis virescens*, cotton budworm; *Helicoverpa zea*, cotton bollworm; *Spodoptera exigua*, beet armyworm;
- 30 *Pectinophora gossypiella*, pink bollworm; *Anthonomus grandis grandis*, boll weevil; *Aphis gossypii*, cotton aphid; *Pseudatomoscelis seriatus*, cotton fleahopper; *Trialeurodes abutilonea*, bandedwinged whitefly; *Lygus lineolaris*, tarnished plant bug; *Melanoplus femurrubrum*,

redlegged grasshopper; *Melanoplus differentialis*, differential grasshopper; *Thrips tabaci*, onion thrips; *Frankliniella fusca*, tobacco thrips; *Tetranychus cinnabarinus*, carmine spider mite; *Tetranychus urticae*, twospotted spider mite; Rice: *Diatraea saccharalis*, sugarcane borer; *Spodoptera frugiperda*, fall armyworm; *Helicoverpa zea*, corn earworm; *Colaspis*
5 *brunnea*, grape colaspis; *Lissorhoptrus oryzophilus*, rice water weevil; *Sitophilus oryzae*, rice weevil; *Nephotettix nigropictus*, rice leafhopper; *Blissus leucopterus leucopterus*, chinch bug; *Acrosternum hilare*, green stink bug; Soybean: *Pseudoplusia includens*, soybean looper; *Anticarsia gemmatilis*, velvetbean caterpillar; *Plathypena scabra*, green cloverworm; *Ostrinia nubilalis*, European corn borer; *Agrotis ipsilon*, black cutworm; *Spodoptera exigua*, beet
10 armyworm; *Heliothis virescens*, cotton budworm; *Helicoverpa zea*, cotton bollworm; *Epilachna varivestis*, Mexican bean beetle; *Myzus persicae*, green peach aphid; *Empoasca fabae*, potato leafhopper; *Acrosternum hilare*, green stink bug; *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus differentialis*, differential grasshopper; *Hylemya platura*, seedcorn maggot; *Sericothrips variabilis*, soybean thrips; *Thrips tabaci*, onion thrips;
15 *Tetranychus turkestani*, strawberry spider mite; *Tetranychus urticae*, twospotted spider mite; Barley: *Ostrinia nubilalis*, European corn borer; *Agrotis ipsilon*, black cutworm; *Schizaphis graminum*, greenbug; *Blissus leucopterus leucopterus*, chinch bug; *Acrosternum hilare*, green stink bug; *Euschistus servus*, brown stink bug; *Delia platura*, seedcorn maggot; *Mayetiola destructor*, Hessian fly; *Petrobia latens*, brown wheat mite; Oil Seed Rape: *Brevicoryne*
20 *brassicae*, cabbage aphid; *Phyllotreta cruciferae*, Flea beetle; *Mamestra configurata*, Bertha armyworm; *Plutella xylostella*, Diamond-back moth; *Delia* ssp., Root maggots.

The sequences of the invention can be used for any application including coating surfaces to target microbes. In this manner, the target microbes include human pathogens or microorganisms. Surfaces that might be coated with the sequences of the invention include
25 carpets and sterile medical facilities. Polymer bound polypeptides of the invention may be used to coat surfaces. Methods for incorporating compositions with antimicrobial properties into polymers are known in the art. See U.S. Patent No.5,847,047, herein incorporated by reference.

30 Definitions

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, amino acid

sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range and include each integer within the defined range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are more fully defined by reference to the specification as a whole.

By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-Beta Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., Persing *et al.*, eds. (1993) *Diagnostic Molecular Microbiology: Principles and Applications*, (American Society for Microbiology, Washington, D.C., 1993). The product of amplification is termed an amplicon.

As used herein, "antisense orientation" includes reference to a duplex polynucleotide sequence, which is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription product is often inhibited.

By "encoding" or "encoded", with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as are present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum*, or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed therein.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For

example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.* (1989) *Nucl. Acids Res.* 17:477-498). Thus, the maize preferred codon for a particular amino acid might be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants is listed in Table 4 of Murray *et al.*, *supra*.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species, or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell, which comprises a heterologous nucleic acid sequence of the invention. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "introduced" in the context of inserting a nucleic acid into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. An "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the nucleic acid molecule or protein as found in its naturally occurring environment. Thus, an isolated or purified nucleic acid molecule or protein is substantially free of other cellular material, or culture medium when produced by recombinant

techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

As used herein, "marker" includes reference to a locus on a chromosome that serves to identify a unique position on the chromosome. A "polymorphic marker" includes reference to a marker, which appears in multiple forms (alleles) such that different forms of the marker, when they are present in a homologous pair, allow transmission of each of the chromosomes of that pair to be followed. Use of one or a plurality of markers may define a genotype.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules, which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular Cloning – A Laboratory Manual* (2nd ed., Vol. 1-3); and Ausubel *et al.*, eds. (1994) *Current Protocols in Molecular Biology* (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.).

As used herein "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where
5 necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. The
10 class of plants, which can be used in the methods of the invention, is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. Preferred plants include, but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. A particularly preferred plant is maize (*Zea mays*).

As used herein, "polynucleotide" includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or
20 a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name
25 just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modification have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including
30 among other things, simple and complex cells.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which

one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but
5 consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide", and "protein" are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a
10 result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Further, this invention contemplates the use of both the
15 methionine containing and the methionine-less amino terminal variants of the protein of the invention.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating
20 transcription in plant cells whether or not its origin is a plant cell. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria which comprise genes expressed in plant cells such as *Agrobacterium* or *Rhizobium*. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, or seeds. Such promoters are
25 referred to as "tissue preferred". A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" or "repressible" promoter is a promoter, which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue preferred, cell type specific, and
30 inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter, which is active under most environmental conditions.

As used herein "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all as a result of deliberate human intervention. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements, which permit transcription of a particular nucleic acid in a host cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The term "residue" or "amino acid residue" or "amino acid" are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass non-natural analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

The term "selectively hybridizes" includes a reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, preferably 90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The terms "stringent conditions" or "stringent hybridization conditions" include reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances.

By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing).

Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides).

Duration of hybridization is generally less than about 24 hours, usually about 4 to 12 hours. Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization for 6 to 8 hours in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a final wash in 0.1X SSC at 60 to 65°C for 30 to 60 minutes.

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, *Anal. Biochem.*, 138:267-284 (1984): $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%CG) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, %CG is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m , hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq 90\%$ identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or

wash at 1, 2, 3, or 4° lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation,
5 hybridization and wash compositions, and desired T_m , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is
10 found in Tijssen, (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", (Elsevier, New York); and Ausubel, *et al.*, Eds., (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing and Wiley-Interscience, New York).

15 As used herein, "transgenic plant" includes reference to a plant, which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell
20 line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events
25 such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, "vector" includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

30 The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison windows", (c) "sequence identity", and (d) "percentage of sequence identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

5 (b) As used herein, "comparison window" means includes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for
10 optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

15 Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2:482; by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol* 48:443; by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444; by computerized implementations
20 of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California, GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp (1988) *Gene* 73:237-244; Higgins and Sharp, (1989) *CABIOS* 5:151-153; Corpet, *et al.*
25 (1988) *Nucleic Acids Research* 16:10881-90; Huang, *et al.* (1992) *Computer Applications in the Biosciences* 8:155-65, and Pearson, *et al.* (1994) *Methods in Molecular Biology* 24:307-331. The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for
30 protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, Ausubel, *et al.*, eds. (1995) *Current*

Protocols in Molecular Biology, Chapter 19, (Greene Publishing and Wiley-Interscience, New York).

GAP uses the algorithm of Needleman and Wunsch (1970) *J Mol Biol* 48:443-453 to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package are 8 and 2, respectively, for protein sequences. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 100. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (*see* Henikoff and Henikoff, *Proc Natl Acad Sci USA* 89:10915).

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters. Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402 or GAP version 10 of Wisconsin Genetic Software Package using default parameters. Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find

longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always <0) .

5 For amino acid sequences, a scoring matrix is used to calculate the cumulative score.

Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X

10 determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA*
15 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an
20 indication of the probability that a match between two nucleotide or two amino acid sequences would occur by chance.

BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences, which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids.
25 Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, (1993) *Comput. Chem.*, 17:149-163) and XNU (Claverie and States (1993) *Comput. Chem.*, 17:191-201) low-complexity filters can be employed alone or in combination.

30 (c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences, which are the same when aligned for maximum correspondence over a specified comparison

window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences, which differ by such conservative substitutions, are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, (1988) *Computer Applic. Biol. Sci.*, 4:11-17 e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

Nucleic Acids

The present invention provides, among other things, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a polynucleotide of the present invention.

A polynucleotide of the present invention is inclusive of:

(a) a polynucleotide encoding a polypeptide of SEQ ID NOS:2, 4, 6, 8, 10, and 12, including exemplary polynucleotides of SEQ ID NOS:1, 3, 5, 7, 9, and 11;

(b) a polynucleotide which is the product of amplification from a *Zea mays* nucleic acid library using primer pairs which selectively hybridize under stringent conditions to loci
5 within a polynucleotide selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, and 11;

(c) a polynucleotide which selectively hybridizes to a polynucleotide of (a) or (b);

(d) a polynucleotide having a specified sequence identity with polynucleotides of (a), (b), or (c);

10 (e) complementary sequences of polynucleotides of (a), (b), (c), or (d);

(f) a polynucleotide comprising at least a specific number of contiguous nucleotides from a polynucleotide of (a), (b), (c), (d), or (e); and

(g) an isolated polynucleotide made by the process of: 1) providing a full-length enriched nucleic acid library, 2) selectively hybridizing the polynucleotide to a polynucleotide
15 of (a), (b), (c), (d), (e), (f), (g), or (h), thereby isolating the polynucleotide from the nucleic acid library.

The present invention provides, among other things, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a polynucleotide of the present invention.

20

A. Polynucleotides Encoding A Polypeptide of the Present Invention

The present invention provides isolated nucleic acids comprising a polynucleotide of the present invention, wherein the polynucleotide encodes a polypeptide of the present invention. Every nucleic acid sequence herein that encodes a polypeptide also, by reference to
25 the genetic code, describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Thus, each silent variation of a nucleic acid, which encodes a polypeptide of the present invention, is implicit in each described
30 polypeptide sequence and is within the scope of the present invention. Accordingly, the present invention includes polynucleotides of the present invention and polynucleotides encoding a polypeptide of the present invention.

B. Polynucleotides Amplified from a Plant Nucleic Acid Library

The present invention provides an isolated nucleic acid comprising a polynucleotide of the present invention, wherein the polynucleotides are amplified, under nucleic acid
5 amplification conditions, from a plant nucleic acid library. Nucleic acid amplification conditions for each of the variety of amplification methods are well known to those of ordinary skill in the art. The plant nucleic acid library can be constructed from a monocot such as a cereal crop. Exemplary cereals include corn, sorghum, alfalfa, canola, wheat, or rice. The plant nucleic acid library can also be constructed from a dicot such as soybean. *Zea*
10 *mays* lines B73, PHRE1, A632, BMS-P2#10, W23, and Mo17 are known and publicly available. Other publicly known and available maize lines can be obtained from the Maize Genetics Cooperation (Urbana, IL). Wheat lines are available from the Wheat Genetics Resource Center (Manhattan, KS).

The nucleic acid library may be a cDNA library, a genomic library, or a library
15 generally constructed from nuclear transcripts at any stage of intron processing. cDNA libraries can be normalized to increase the representation of relatively rare cDNAs. In optional embodiments, the cDNA library is constructed using an enriched full-length cDNA synthesis method. Examples of such methods include Oligo-Capping (Maruyama and Sugano (1994) *S. Gene* 138:171-174), Biotinylated CAP Trapper (Carninci, *et al.* (1996) *Genomics*
20 37:327-336), and CAP Retention Procedure (Edery *et al.* (1995) *Molecular and Cellular Biology* 15:3363-3371). Rapidly growing tissues or rapidly dividing cells are preferred for use as an mRNA source for construction of a cDNA library. Growth stages of corn is described in "How a Corn Plant Develops," Special Report No. 48, Iowa State University of Science and Technology Cooperative Extension Service, Ames, Iowa, Reprinted February
25 1993.

A polynucleotide of this embodiment (or subsequences thereof) can be obtained, for example, by using amplification primers which are selectively hybridized and primer extended, under nucleic acid amplification conditions, to at least two sites within a polynucleotide of the present invention, or to two sites within the nucleic acid which flank and
30 comprise a polynucleotide of the present invention, or to a site within a polynucleotide of the present invention and a site within the nucleic acid which comprises it. Methods for obtaining 5' and/or 3' ends of a vector insert are well known in the art. See, e.g., RACE (Rapid

Amplification of Complementary Ends) as described in Frohman, M. A., M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White, eds. (1990) *PCR Protocols: A Guide to Methods and Applications*, pp. 28-38 (Academic Press, Inc., San Diego)); see also, U.S. Pat. No. 5,470,722, and Ausubel, *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Unit 15.6 (Greene Publishing and Wiley-Interscience, New York); Frohman and Martin (1989) *Techniques* 1:165.

Optionally, the primers are complementary to a subsequence of the target nucleic acid which they amplify but may have a sequence identity ranging from about 85% to 99% relative to the polynucleotide sequence which they are designed to anneal to. As those skilled in the art will appreciate, the sites to which the primer pairs will selectively hybridize are chosen such that a single contiguous nucleic acid can be formed under the desired nucleic acid amplification conditions. The primer length in nucleotides is selected from the group of integers consisting of from at least 15 to 50. Thus, the primers can be at least 15, 18, 20, 25, 30, 40, or 50 nucleotides in length. Those of skill will recognize that a lengthened primer sequence can be employed to increase specificity of binding (i.e., annealing) to a target sequence. A non-annealing sequence at the 5' end of a primer (a "tail") can be added, for example, to introduce a cloning site at the terminal ends of the amplicon.

The amplification products can be translated using expression systems well known to those of skill in the art. The resulting translation products can be confirmed as polypeptides of the present invention by, for example, assaying for the appropriate catalytic activity (e.g., specific activity and/or substrate specificity), or verifying the presence of one or more linear epitopes, which are specific to a polypeptide of the present invention. Methods for protein synthesis from PCR derived templates are known in the art and available commercially. See, e.g., Amersham Life Sciences, Inc, Catalog '97, p.354.

C. Polynucleotides Which Selectively Hybridize to a Polynucleotide of (A) or (B)

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides selectively hybridize, under selective hybridization conditions, to a polynucleotide of section (A) or (B) as discussed above. Such sequences may encode polypeptides that retain the biological activity of the disclosed sequences. Thus, the polynucleotides of this embodiment can be used for isolating, detecting, and/or quantifying nucleic acids comprising the polynucleotides of (A) or (B). For example,

polynucleotides of the present invention can be used to identify, isolate, or amplify partial or full-length clones in a deposited library. In some embodiments, the polynucleotides are genomic or cDNA sequences isolated or otherwise complementary to a cDNA from a dicot or monocot nucleic acid library. Exemplary species of monocots and dicots include, but are not limited to: maize, canola, soybean, cotton, wheat, sorghum, sunflower, alfalfa, oats, sugar cane, millet, barley, and rice. The cDNA library comprises at least 50% to 95% full-length sequences (for example, at least 50%, 60%, 70%, 80%, 90%, or 95% full-length sequences). The cDNA libraries can be normalized to increase the representation of rare sequences. See, e.g., U.S. Patent No. 5,482,845. Low stringency hybridization conditions are typically, but not exclusively, employed with sequences having a reduced sequence identity relative to complementary sequences. Moderate and high stringency conditions can optionally be employed for sequences of greater identity. Low stringency conditions allow selective hybridization of sequences having about 70% to 80% sequence identity and can be employed to identify orthologous or paralogous sequences.

15

D. Polynucleotides Having a Specific Sequence Identity with the Polynucleotides of (A), (B) or (C)

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides have a specified identity at the nucleotide level to a polynucleotide as disclosed above in sections (A), (B), or (C), above. Such sequences may encode polypeptides that retain biological activity of the disclosed sequences. Identity can be calculated using, for example, the BLAST or GAP algorithms under default conditions. The percentage of identity to a reference sequence is at least 60% and, rounded upwards to the nearest integer, can be expressed as an integer selected from the group of integers consisting of from 60 to 99. Thus, for example, the percentage of identity to a reference sequence can be at least 70%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%.

30

E. Polynucleotides Encoding a Protein Having a Subsequence from a Prototype Polypeptide and Cross-Reactive to the Prototype Polypeptide

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides encode a protein having a subsequence of
5 contiguous amino acids from a prototype polypeptide of the present invention such as are provided in section (A), above. The subsequences of a nucleotide sequence may encode protein fragments that retain the biological activity of the native protein and hence modulate disease resistance. Alternatively, subsequences of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity.
10 Thus, subsequences of a nucleotide sequence may range from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, and up to the full-length nucleotide sequence encoding the proteins of the invention.

The length of contiguous amino acids from the prototype polypeptide is selected from the group of integers consisting of from at least 10 to the number of amino acids within the
15 prototype sequence. Thus, for example, the polynucleotide can encode a polypeptide having a biologically active subsequence having at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240 or more contiguous amino acids from the prototype polypeptide. Further, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 20,
20 such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 25, 50, 100, or 200 nucleotides.

Thus, a subsequence of a sequence of a nucleotide sequence of the invention may encode a biologically active portion of an encoded protein, or it may be a fragment that can be
25 used as a hybridization probe or PCR primer using methods disclosed below. A biologically active portion of a protein of the invention can be prepared by isolating a portion of one of the nucleotide sequences of the invention, expressing the encoded portion of the protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the protein. Nucleic acid molecules that are subsequences of a nucleotide sequence of the
30 invention comprise at least 16, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1,000, 1,100, 1,200, 1,300, or 1,396 nucleotides, or up to the number of nucleotides present in a full-length of the nucleotide sequences disclosed herein (for

example, 459, 597, 1137, 830, 445, and 1397 nucleotides for SEQ ID NOS:1, 3, 5, 7, 9, and 11, respectively).

The proteins encoded by polynucleotides of this embodiment, when presented as an immunogen, elicit the production of polyclonal antibodies which specifically bind to a
5 prototype polypeptide such as (but not limited to) a polypeptide encoded by the polynucleotide of sections (A) or (B) above. Generally, however, a protein encoded by a polynucleotide of this embodiment does not bind to antisera raised against the prototype polypeptide when the antisera has been fully immunosorbed with the prototype polypeptide. Methods of making and assaying for antibody binding specificity/affinity are well known in
10 the art. Exemplary immunoassay formats include ELISA, competitive immunoassays, radioimmunoassays, Western blots, indirect immunofluorescent assays and the like.

In a preferred assay method, fully immunosorbed and pooled antisera that is elicited to the prototype polypeptide can be used in a competitive binding assay to test the protein. The concentration of the prototype polypeptide required to inhibit 50% of the binding of the
15 antisera to the prototype polypeptide is determined. If the amount of the protein required to inhibit binding is less than twice the amount of the prototype protein, then the protein is said to specifically bind to the antisera elicited to the immunogen. Accordingly, the proteins of the present invention embrace allelic variants, conservatively modified variants, and minor recombinant modifications to a prototype polypeptide.

20 A polynucleotide of the present invention optionally encodes a protein having a molecular weight of the non-glycosylated protein within 20% of the molecular weight of the full-length non-glycosylated polypeptides of the present invention. Molecular weight can be readily determined by SDS-PAGE under reducing conditions. Optionally, the molecular weight is within 15% of a full-length polypeptide of the present invention, more preferably
25 within 10% or 5%, and most preferably within 3%, 2%, or 1% of a full-length polypeptide of the present invention.

Optionally, the polynucleotides of this embodiment will encode a protein having a specific enzymatic activity at least 50%, 60%, 70%, 80%, or 90% of a cellular extract comprising the native, endogenous full-length polypeptide of the present invention. Further,
30 the proteins encoded by polynucleotides of this embodiment will optionally have a substantially similar affinity constant (K_m) and/or catalytic activity (*i.e.*, the microscopic rate constant, k_{cat}) as the native endogenous, full-length protein. Those of skill in the art will

recognize that k_{cat}/K_m value determines the specificity for competing substrates and is often referred to as the specificity constant. Proteins of this embodiment can have a k_{cat}/K_m value at least 10% of a full-length polypeptide of the present invention as determined using the endogenous substrate of that polypeptide. Optionally, the k_{cat}/K_m value will be at least 20%,
5 30%, 40%, 50%, and most preferably at least 60%, 70%, 80%, 90%, or 95% the k_{cat}/K_m value of the full-length polypeptide of the present invention. Determination of k_{cat} , K_m , and k_{cat}/K_m can be determined by any number of means well known to those of skill in the art. For example, the initial rates (*i.e.*, the first 5% or less of the reaction) can be determined using rapid mixing and sampling techniques (*e.g.*, continuous-flow, stopped-flow, or rapid
10 quenching techniques), flash photolysis, or relaxation methods (*e.g.*, temperature jumps) in conjunction with such exemplary methods of measuring as spectrophotometry, spectrofluorimetry, nuclear magnetic resonance, or radioactive procedures. Kinetic values are conveniently obtained using a Lineweaver-Burk or Eadie-Hofstee plot.

15 *F. Polynucleotides Complementary to the Polynucleotides of (A)-(E)*

The present invention provides isolated nucleic acids comprising polynucleotides complementary to the polynucleotides of sections A-E, above. As those of skill in the art will recognize, complementary sequences base pair throughout the entirety of their length with the polynucleotides of sections (A)-(E) (*i.e.*, have 100% sequence identity over their entire
20 length). Complementary bases associate through hydrogen bonding in double stranded nucleic acids. For example, the following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.

G. Polynucleotides that are Subsequences of the Polynucleotides of (A)-(F)

25 The present invention provides isolated nucleic acids comprising polynucleotides which comprise at least 15 contiguous bases from the polynucleotides of sections (A) (B), (C), (D), (E), or (F) (*i.e.*, sections (A)-(F), as discussed above). A subsequence of a nucleotide sequence of the invention may encode a biologically active portion of a protein, or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed
30 elsewhere herein. Subsequences of a nucleotide sequence of the invention that are useful as hybridization probes or PCR primers generally need not encode a biologically active portion of a protein.

The length of the polynucleotide is given as an integer selected from the group consisting of from at least 15 to the length of the nucleic acid sequence from which the polynucleotide is a subsequence of. Thus, for example, polynucleotides of the present invention are inclusive of polynucleotides comprising at least 15, 16, 17, 18, 19, 20, 21, 22,
5 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47,
48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72,
73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190,
200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 350, 400, 450, 500, 550, 600, 650, 700,
750, 800, 850, 900, 950, or 1000 contiguous nucleotides in length from the polynucleotides of
10 sections (A) through (F). Optionally, the number of such subsequences encoded by a
polynucleotide of the instant embodiment can be any integer selected from the group
consisting of from 1 to 1000, such as 2, 3, 4, or 5. The subsequences can be separated by any
integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5,
10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000
15 nucleotides.

Subsequences can be made by *in vitro* synthetic, *in vitro* biosynthetic, or *in vivo* recombinant methods. In optional embodiments, subsequences can be made by nucleic acid amplification. For example, nucleic acid primers will be constructed to selectively hybridize to a sequence (or its complement) within, or co-extensive with, the coding region.

20 The subsequences of the present invention can comprise structural characteristics of the sequence from which it is derived. Alternatively, the subsequences can lack certain structural characteristics of the larger sequence from which it is derived such as a poly (A) tail. Optionally, a subsequence from a polynucleotide encoding a polypeptide having at least one linear epitope in common with a prototype polypeptide sequence as provided in (a),
25 above, may encode an epitope in common with the prototype sequence. Alternatively, the subsequence may not encode an epitope in common with the prototype sequence but can be used to isolate the larger sequence by, for example, nucleic acid hybridization with the sequence from which it is derived. Subsequences can be used to modulate or detect gene expression by introducing into the subsequences compounds which bind, intercalate, cleave
30 and/or crosslink to nucleic acids. Exemplary compounds include acridine, psoralen, phenanthroline, naphthoquinone, daunomycin or chloroethylaminoaryl conjugates.

H. Polynucleotides that are Variants of the Polynucleotides of (A)-(G).

By "variants" is intended substantially similar sequences. For nucleotide sequences, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of one of the polypeptides of the invention. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis, but which still encode a protein of the invention. Generally, variants of a particular nucleotide sequence of the invention will have at least about 40%, 50%, 60%, 65%, 70%, generally at least about 75%, 80%, 85%, preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity to that particular nucleotide sequence as determined by sequence alignment programs described elsewhere herein using default parameters.

15

I. Polynucleotides from a Full-length Enriched cDNA Library Having the Physico-Chemical Property of Selectively Hybridizing to a Polynucleotide of (A)-(H)

The present invention provides an isolated polynucleotide from a full-length enriched cDNA library having the physico-chemical property of selectively hybridizing to a polynucleotide of sections (A), (B), (C), (D), (E), (F), (G), or (H) as discussed above. Methods of constructing full-length enriched cDNA libraries are known in the art and discussed briefly below. The cDNA library comprises at least 50% to 95% full-length sequences (for example, at least 50%, 60%, 70%, 80%, 90%, or 95% full-length sequences). The cDNA library can be constructed from a variety of tissues from a monocot or dicot at a variety of developmental stages. Exemplary species include maize, wheat, rice, canola, soybean, cotton, sorghum, sunflower, alfalfa, oats, sugar cane, millet, barley, and rice. Methods of selectively hybridizing, under selective hybridization conditions, a polynucleotide from a full-length enriched library to a polynucleotide of the present invention are known to those of ordinary skill in the art. Any number of stringency conditions can be employed to allow for selective hybridization. In optional embodiments, the stringency allows for selective hybridization of sequences having at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, up to 100% sequence identity over the length of the hybridized

30

region. Full-length enriched cDNA libraries can be normalized to increase the representation of rare sequences.

J. Polynucleotide Products Made by a cDNA Isolation Process

5 The present invention provides an isolated polynucleotide made by the process of: 1) providing a full-length enriched nucleic acid library; and 2) selectively hybridizing the polynucleotide to a polynucleotide of sections (A), (B), (C), (D), (E), (F), (G), (H), or (I) as discussed above, and thereby isolating the polynucleotide from the nucleic acid library. Full-length enriched nucleic acid libraries are constructed and selective hybridization conditions
10 are used, as discussed below. Such techniques, as well as nucleic acid purification procedures, are well known in the art. Purification can be conveniently accomplished using solid-phase methods; such methods are well known to those of skill in the art and kits are available from commercial suppliers such as Advanced Biotechnologies (Surrey, UK). For example, a polynucleotide of sections (A)-(H) can be immobilized to a solid support such as a
15 membrane, bead, or particle. See, *e.g.*, U.S. Patent No. 5,667,976. The polynucleotide product of the present process is selectively hybridized to an immobilized polynucleotide and the solid support is subsequently isolated from non-hybridized polynucleotides by methods including, but not limited to, centrifugation, magnetic separation, filtration, electrophoresis, and the like.

20

Construction of Nucleic Acids

 The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some
25 embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a monocot.

 The nucleic acids may conveniently comprise sequences in addition to a polynucleotide of the present invention. For example, a multi-cloning site comprising one or more endonuclease restriction sites may be inserted into the nucleic acid to aid in isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the
30 translated polynucleotide of the present invention. For example, a hexa-histidine marker sequence provides a convenient means to purify the proteins of the present invention. A polynucleotide of the present invention can be attached to a vector, adapter, or linker for

cloning and/or expression of a polynucleotide of the present invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Typically, the length of a nucleic acid of the present invention less the length of its polynucleotide of the present invention is less than 20 kilobase pairs, often less than 15 kb, and frequently less than 10 kb. Use of cloning vectors, expression vectors, adapters, and linkers is well known and extensively described in the art. For a description of various nucleic acids see, for example, Stratagene Cloning Systems, Catalogs 1999 (La Jolla, CA); and, Amersham Life Sciences, Inc, Catalog '99 (Arlington Heights, IL).

A. Recombinant Methods for Constructing Nucleic Acids

The isolated nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art. In some embodiments, oligonucleotide probes, which selectively hybridize, under stringent conditions, to the polynucleotides of the present invention are used to identify the desired sequence in a cDNA or genomic DNA library. Isolation of RNA and construction of cDNA and genomic libraries is well known to those of ordinary skill in the art. See, e.g., Clark, ed. (1997) *Plant Molecular Biology: A Laboratory Manual* (Springer-Verlag, Berlin); and, Ausubel, *et al.*, eds. (1995) *Current Protocols in Molecular Biology* (Greene Publishing and Wiley-Interscience, New York).

A1. Full-length Enriched cDNA Libraries

A number of cDNA synthesis protocols have been described which provide enriched full-length cDNA libraries. Enriched full-length cDNA libraries are constructed to comprise at least 60%, and more preferably at least 70%, 80%, 90% or 95% full-length inserts amongst clones containing inserts. The length of insert in such libraries can be at least 2, 3, 4, 5, 6, 7, 8, 9, 10 or more kilobase pairs. Vectors to accommodate inserts of these sizes are known in the art and available commercially. See, e.g., Stratagene's lambda ZAP Express (cDNA cloning vector with 0 to 12 kb cloning capacity). An exemplary method of constructing a greater than 95% pure full-length cDNA library is described by Carninci *et al.* (1996) *Genomics* 37:327-

336. Other methods for producing full-length libraries are known in the art. See, e.g., Edery *et al.* (1995) *Mol. Cell Biol.* 15(6):3363-3371; and, PCT Application WO 96/34981.

A2. Normalized or Subtracted cDNA Libraries

5 A non-normalized cDNA library represents the mRNA population of the tissue it was made from. Since unique clones are out-numbered by clones derived from highly expressed genes their isolation can be laborious. Normalization of a cDNA library is the process of creating a library in which each clone is more equally represented. Construction of normalized libraries is described in Ko (1990) *Nucl. Acids. Res.* 18(19):5705-5711; Patanjali *et al.* (1991) *Proc. Natl. Acad. U.S.A.* 88:1943-1947; U.S. Patents 5,482,685, 5,482,845, and 5,637,685. In an exemplary method described by Soares *et al.*, normalization resulted in reduction of the abundance of clones from a range of four orders of magnitude to a narrow range of only 1 order of magnitude. *Proc. Natl. Acad. Sci. USA*, 91:9228-9232 (1994).

15 Subtracted cDNA libraries are another means to increase the proportion of less abundant cDNA species. In this procedure, cDNA prepared from one pool of mRNA is depleted of sequences present in a second pool of mRNA by hybridization. The cDNA:mRNA hybrids are removed and the remaining un-hybridized cDNA pool is enriched for sequences unique to that pool. See, Foote *et al.*, Clark, ed. (1997) *Plant Molecular Biology: A Laboratory Manual* (Springer-Verlag, Berlin); Kho and Zarbl (1991) *Technique*, 3(2):58-63; Sive and St. John (1988) *Nucl. Acids Res.*, 16(22):10937; Ausubel, *et al.*, eds. (1995) *Current Protocols in Molecular Biology* (Greene Publishing and Wiley-Interscience, New York; and, Swaroop *et al.* (1991) *Nucl. Acids Res.*, 19(17):4725-4730. cDNA subtraction kits are commercially available. See, e.g., PCR-Select (Clontech, Palo Alto, CA).

25 To construct genomic libraries, large segments of genomic DNA are generated by fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. Methodologies to accomplish these ends, and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate molecular biological techniques and instructions sufficient to direct persons of skill through many construction, cloning, and screening methodologies are found in Sambrook, *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Vols. 1-3 (2nd ed., Cold Spring Harbor Laboratory), Berger and Kimmel, eds. (1987) "Methods in Enzymology," Vol. 152: *Guide to Molecular Cloning Techniques* (San Diego: Academic

30

Press, Inc.), Ausubel, *et al.*, eds. (1995) *Current Protocols in Molecular Biology* (Greene Publishing and Wiley-Interscience, New York 1995); Clark, ed. (1997) *Plant Molecular Biology: A Laboratory Manual*, (Springer-Verlag, Berlin). Kits for construction of genomic libraries are also commercially available.

5 The cDNA or genomic library can be screened using a probe based upon the sequence of a polynucleotide of the present invention such as those disclosed herein. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the
10 wash medium can be stringent.

 The nucleic acids of interest can also be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of polynucleotides of the present invention and related genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods
15 may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. The T4 gene 32 protein (Boehringer Mannheim) can be used to improve yield of long PCR products.

 PCR-based screening methods have been described. Wilfinger *et al.* describe a PCR-
20 based method in which the longest cDNA is identified in the first step so that incomplete clones can be eliminated from study. *BioTechniques*, 22(3): 481-486 (1997). Such methods are particularly effective in combination with a full-length cDNA construction methodology, above.

25 B. Synthetic Methods for Constructing Nucleic Acids

 The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68:90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68:109-151; the diethylphosphoramidite method of Beaucage *et al.*, (1981) *Tetra. Lett.*
30 22:1859-1862; the solid phase phosphoramidite triester method described by Beaucage and Caruthers, (1981) *Tetra. Letts.* 22(20):1859-1862, *e.g.*, using an automated synthesizer, *e.g.*, as described in Needham-VanDevanter *et al.*, *Nucleic Acids Res.* (1984) 12:6159-6168; and,

the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill will recognize that while
5 chemical synthesis of DNA is best employed for sequences of about 100 bases or less, longer sequences may be obtained by the ligation of shorter sequences.

Recombinant Expression Cassettes

The present invention further provides recombinant expression cassettes comprising a
10 nucleic acid of the present invention. A nucleic acid sequence coding for the desired polynucleotide of the present invention, for example a cDNA or a genomic sequence encoding a full length polypeptide of the present invention, can be used to construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a polynucleotide of the present invention operably
15 linked to transcriptional initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

For example, plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g.,
20 one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

A number of promoters can be used in the practice of the invention. A plant promoter
25 fragment can be employed which will direct expression of a polynucleotide of the present invention in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and stages of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the ubiquitin 1 promoter (Christensen, *et al.* (1992) *Plant Mol Biol* 18:675-689; Bruce, *et al.* 1989) *Proc Natl Acad Sci USA* 86:9692-9696), the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Patent No,
30

5,683,439), the *Nos* promoter, the pEmu promoter, the rubisco promoter, the GRP1-8 promoter, the maize constitutive promoters described in PCT Publication No. WO 99/43797 which include the histone H2B, metallothionein, alpha-tubulin 3, elongation factor efla, ribosomal protein rps8, chlorophyll a/b binding protein, and glyceraldehyde-3-phosphate dehydrogenase promoters, and other transcription initiation regions from various plant genes known to those of skill.

Where low level expression is desired, weak promoters will be used. It is recognized that weak inducible promoters may be used. Additionally, either a weak constitutive or a weak tissue specific promoter may be used. Generally, by a "weak promoter" is intended a promoter that drives expression of a coding sequence at a low level. By low level is intended at levels of about 1/1000 transcripts to about 1/100,000 transcripts to about 1/500,000 transcripts. Alternatively, it is recognized that weak promoters also encompass promoters that are expressed in only a few cells and not in others to give a total low level of expression. Such weak constitutive promoters include, for example, the core promoter of the Rsyn7 (PCT Publication No. WO 97/44756), the core 35S CaMV promoter, and the like. Where a promoter is expressed at unacceptably high levels, portions of the promoter sequence can be deleted or modified to decrease expression levels. Additionally, to obtain a varied series in the level of expression, one can also make a set of transgenic plants containing the polynucleotides of the present invention with a strong constitutive promoter, and then rank the transgenic plants according to the observed level of expression. The transgenic plants will show a variety in performance, from high expression to low expression. Factors such as chromosomal position effect, cosuppression, and the like will affect the expression of the polynucleotide.

Alternatively, the plant promoter can direct expression of a polynucleotide of the present invention under environmental control. Such promoters are referred to here as "inducible" promoters. Environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light. Examples of inducible promoters are the *Adh1* promoter, which is inducible by hypoxia or cold stress, the *Hsp70* promoter, which is inducible by heat stress, and the *PPDK* promoter, which is inducible by light. Examples of pathogen-inducible promoters include those from proteins, which are induced following infection by a pathogen; *e.g.*, PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, etc. See, for example, Redolfi, *et al.* (1983) *Meth J. Plant Pathol.*

89:245-254; Uknes *et al.* (1992) *The Plant Cell* 4:645-656; Van Loon (1985) *Plant Mol. Virol.* 4:111-116; PCT Publication No. WO 99/43819.

Of interest are promoters that are expressed locally at or near the site of pathogen infection. See, for example, Marineau, *et al.* (1987) *Plant Mol Biol* 9:335-342; Matton *et al.* (1987) *Molecular Plant-Microbe Interactions* 2:325-342; Somssich *et al.* (1986) *Proc Natl Acad Sci USA* 83:2427-2430; Somssich *et al.* (1988) *Mol Gen Genetics* 2:93-98; Yang, *Proc Natl Acad Sci USA* 93:14972-14977. See also, Chen, *et al.* (1996) *Plant J* 10:955-966; Zhang and Sing (1994) *Proc Natl Acad Sci USA* 91:2507-2511; Warner, *et al.* (1993) *Plant J* 3:191-201, and Siebertz, *et al.* (1989) *Plant Cell* 1:961-968, all of which are herein incorporated by reference. Of particular interest is the inducible promoter for the maize PRms gene, whose expression is induced by the pathogen *Fusarium moniliforme* (see, for example, Cordero, *et al.* (1992) *Physiol Molec Plant Path* 41:189-200 and is herein incorporated by reference.

Additionally, as pathogens find entry into plants through wounds or insect damage, a wound inducible promoter may be used in the constructs of the invention. Such wound inducible promoter include potato proteinase inhibitor (pin II) gene (Ryan (1990) *Annu Rev Phytopath* 28:425-449; Duan, *et al.* (1996) *Nat Biotech* 14:494-498); wun1 and wun 2, US Patent No. 5,428,148; win1 and win2 (Stanford *et al.* (1989) *Mol Gen Genet* 215:200-208); systemin (McGurl, *et al.* (1992) *Science* 225:1570-1573); WIP1 (Rohmeier, *et al.* (1993) *Plant Mol Biol* 22:783-792; Eckelkamp, *et al.* (1993) *FEB Letters* 323:73-76); MPI gene (Cordero, *et al.* (1994) *The Plant J* 6(2):141-150); and the like, herein incorporated by reference.

Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds, or flowers. Exemplary promoters include the anther specific promoter 5126 (U.S. Patent Nos. 5,689,049 and 5,689,051), glob-1 promoter, and gamma-zein promoter. An exemplary promoter for leaf- and stalk-preferred expression is MS8-15 (PCT Publication no. WO 98/00533). Examples of seed-preferred promoters included, but are not limited to, 27 kD gamma zein promoter and waxy promoter (Boronat, *et al.* (1986) *Plant Sci*, 47:95-102; Reina, *et al.* (1990) *Nucleic Acids Res* 18(21):6426; and Kloesgen, *et al.* (1986) *Mol Gen Genet* 203:237-244). Promoters that express in the embryo, pericarp, and endosperm are disclosed in WO 00/11177 and WO 00/12733, both of which are hereby incorporated by reference. The operation of a promoter may also vary depending on its location in the genome. Thus, a

developmentally regulated promoter may become fully or partially constitutive in certain locations. A developmentally regulated promoter can also be modified, if necessary, for weak expression.

Both heterologous and non-heterologous (i.e. endogenous) promoters can be employed to direct expression of the nucleic acids of the present invention. These promoters can also be used, for example, in recombinant expression cassettes to drive expression of antisense nucleic acids to reduce, increase, or alter concentration and/or composition of the proteins of the present invention in a desired tissue. Thus, in some embodiments, the nucleic acid construct will comprise a promoter functional in a plant cell, such as in *Zea Mays*, operably linked to a polynucleotide of the present invention. Promoters useful in these embodiments include the endogenous promoters driving expression of a polypeptide of the present invention.

In some embodiments, isolated nucleic acids which serve as promoter or enhancer elements can be introduced in the appropriate position (generally upstream) of a non-heterologous form of a polynucleotide of the present invention so as to up or down regulate expression of a polynucleotide of the present invention. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868), or isolated promoters can be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene. Gene expression can be modulated under conditions suitable for plant growth so as to alter the total concentration and/or alter the composition of the polypeptides of the present invention in plant cell. Thus, the present invention provides compositions, and methods for making, heterologous promoters and/or enhancers operably linked to a native, endogenous (i.e., non-heterologous) form of a polynucleotide of the present invention.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold, Buchman and Berg (1988) *Mol. Cell Biol.* 8:4395-4405; Callis *et al.* (1987) *Genes Dev.* 1: 1183-1200. Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, Freeling and Walbot, eds. (1994) *The Maize Handbook*, Chapter 116 (Springer, New York).

The vector comprising the sequences from a polynucleotide of the present invention will typically comprise a marker gene, which confers a selectable phenotype on plant cells. Usually, the selectable marker gene will encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g., the *aada* gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance, genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-induced (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.* (1987) *Meth. In Enzymol.* 153:253-277. These vectors are plant integrating vectors in that upon transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.* (1987) *Gene*, 61:1-11 and Berger *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:8402-8406. Another useful vector herein is plasmid pBI101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA).

A polynucleotide of the present invention can be expressed in either sense or anti-sense orientation as desired. It will be appreciated that control of gene expression in either sense or anti-sense orientation can have a direct impact on the observable plant characteristics. Antisense technology can be conveniently used to inhibit gene expression in plants. To
5 accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the anti-sense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been shown that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy *et al* (1988) *Proc. Nat'l.*
10 *Acad. Sci (USA)* 85:8805-8809; and Hiatt *et al.* U.S. Patent No. 4,801,340.

Another method of suppression is sense suppression. Introduction of nucleic acid configured in the sense orientation has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli *et al.* (1990) *The Plant Cell* 2:279-289 and U.S.
15 Patent No. 5,034,323.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of plant genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered,
20 and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. The design and use of target RNA-specific ribozymes is described in Haseloff *et al.* (1988) *Nature* 334:585-591.

A variety of cross-linking agents, alkylating agents and radical generating species as
25 pendant groups on polynucleotides of the present invention can be used to bind, label, detect, and/or cleave nucleic acids. For example, Vlassov, V.V., *et al.* (1986) *Nucleic Acids Res* 14:4065-4076, describe covalent bonding of a single-stranded DNA fragment with alkylating derivatives of nucleotides complementary to target sequences. A report of similar work by the same group is that by Knorre, *et al.* (1985) *Biochimie* 67:785-789. Iverson and Dervan also
30 showed sequence-specific cleavage of single-stranded DNA mediated by incorporation of a modified nucleotide which was capable of activating cleavage (*J Am Chem Soc* (1987) 109:1241-1243). Meyer, R.B. *et al.* (1989) *J Am Chem Soc* 111:8517-8519, effect covalent

crosslinking to a target nucleotide using an alkylating agent complementary to the single-stranded target nucleotide sequence. A photoactivated crosslinking to single-stranded oligonucleotides mediated by psoralen was disclosed by Lee, *et al.* (1988) *Biochemistry* 27:3197-3203. Use of crosslinking in triple-helix forming probes was also disclosed by Home
5 *et al.* (1990) *J Am Chem Soc* 112:2435-2437. Use of N4, N4-ethanocytosine as an alkylating agent to crosslink to single-stranded oligonucleotides has also been described by Webb and Matteucci (1986) *J Am Chem Soc* 108:2764-2765; *Nucleic Acids Res* (1986) 14:7661-7674; Feteritz *et al.* (1991) *J. Am. Chem. Soc.* 113:4000. Various compounds to bind, detect, label, and/or cleave nucleic acids are known in the art. See, for example, U.S. Patent Nos.
10 5,543,507; 5,672,593; 5,484,908; 5,256,648; and 5,681,941.

Polypeptides

The isolated proteins of the present invention comprise a polypeptide having at least 10 amino acids encoded by any one of the polynucleotides of the present invention as
15 discussed more fully, above, or polypeptides which are conservatively modified variants thereof. The proteins of the present invention or variants thereof can comprise any number of contiguous amino acid residues from a polypeptide of the present invention, wherein that number is selected from the group of integers consisting of from 10 to the number of residues in a full-length polypeptide of the present invention. Optionally, this subsequence of
20 contiguous amino acids is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, or 40 amino acids in length, often at least 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acids in length.

By "variant" protein is intended a protein derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-
25 terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Variant proteins encompassed by the present invention are biologically active, that is they continue to possess the desired biological activity of the native protein, that is, modulate disease resistance as described herein. Such variants may result
30 from, for example, genetic polymorphism or from human manipulation. Biologically active variants of a native protein of the invention will have at least about 40%, 50%, 60%, 65%, 70%, generally at least about 75%, 80%, 85%, preferably at least about 90%, 91%, 92%, 93%,

94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity to the amino acid sequence for the native protein as determined by sequence alignment programs described elsewhere herein using default parameters. A biologically active variant of a protein of the invention may differ from that protein by as few as 1-15 amino acid residues, as few as 1-10, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

As those of skill will appreciate, the present invention includes catalytically active polypeptides of the present invention (*i.e.*, enzymes). Catalytically active polypeptides have a specific activity of at least 20%, 30%, or 40%, and preferably at least 50%, 60%, or 70%, and most preferably at least 80%, 90%, or 95% that of the native (non-synthetic), endogenous polypeptide. Further, the substrate specificity (k_{cat}/K_m) is optionally substantially similar to the native (non-synthetic), endogenous polypeptide. Typically, the K_m will be at least 30%, 40%, or 50%, that of the native (non-synthetic), endogenous polypeptide; and more preferably at least 60%, 70%, 80%, or 90%. Methods of assaying and quantifying measures of enzymatic activity and substrate specificity (k_{cat}/K_m), are well known to those of skill in the art.

Generally, the proteins of the present invention will, when presented as an immunogen, elicit production of an antibody specifically reactive to a polypeptide of the present invention. Further, the proteins of the present invention will not bind to antisera raised against a polypeptide of the present invention which has been fully immunosorbed with the same polypeptide. Immunoassays for determining binding are well known to those of skill in the art. A preferred immunoassay is a competitive immunoassay as discussed *infra*. Thus, the proteins of the present invention can be employed as immunogens for constructing antibodies immunoreactive to a protein of the present invention for such exemplary utilities as immunoassays or protein purification techniques.

The proteins of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the proteins can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan

Publishing Company, New York) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative
5 substitutions, such as exchanging one amino acid with another having similar properties, may be preferable.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the proteins of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof.
10 Such variants will continue to possess the ability to modulate disease resistance. Obviously, the mutations that will be made in the DNA encoding the variant must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

The deletions, insertions, and substitutions of the protein sequences encompassed
15 herein are not expected to produce radical changes in the characteristics of the protein. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays, as described elsewhere herein.

As discussed elsewhere herein, variant nucleotide sequences and proteins also
20 encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different coding sequences can be manipulated to create a new polypeptide possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can
25 be homologously recombined *in vitro* or *in vivo*.

Expression of Proteins in Host Cells

Using the nucleic acids of the present invention, one may express a protein of the present invention in a recombinantly engineered cell such as bacteria, yeast, insect,
30 mammalian, or preferably plant cells. The cells produce the protein in a non-natural condition. (e.g., in quantity, composition, location, and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of a nucleic acid encoding a protein of the present invention. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

5 Briefly, the expression of isolated nucleic acids encoding a protein of the present invention will typically be achieved by operably linking, for example, the DNA or cDNA to a promoter (which is either constitutive or regulatable), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation
10 terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding a protein of the present invention. To obtain high level expression of a cloned gene, it is desirable to construct expression vectors which contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. One of skill would recognize that modifications could be
15 made to a protein of the present invention without diminishing its biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create
20 conveniently located purification sequences. Restriction sites or termination codons can also be introduced.

A. Expression in Prokaryotes

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are
25 represented by various strains of *E. coli*; however, other microbial strains may also be used. Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding sequences, include such commonly used promoters as the beta lactamase (penicillinase) and lactose (lac) promoter systems (Chang *et al.* (1997) *Nature* 198:1056), the tryptophan (trp)
30 promoter system (Goeddel *et al.* (1980) *Nucleic Acids Res.* 8:4057) and the lambda derived P L promoter and N-gene ribosome binding site (Shimatake *et al.* (1981) *Nature* 292:128). The inclusion of selection markers in DNA vectors transfected in *E. coli* is also useful.

Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

5 The vector is selected to allow introduction into the appropriate host cell. Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA. Expression systems for expressing a protein of the present invention are available using *Bacillus sp.* and *Salmonella* (Palva *et al.* (1983) *Gene* 22:229-235; Mosbach, *et al.* (1983) *Nature* 302:543-545).

10 B. Expression in Eukaryotes

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, a polynucleotide of the present invention can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as
15 expression systems for production of the proteins of the instant invention.

Synthesis of heterologous proteins in yeast is well known. Sherman, *et al.* (1982) *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory) is a well recognized work describing the various methods available to produce the protein in yeast. Two widely utilized yeasts for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia*
20 *pastoris*. Vectors, strains, and protocols for expression in *Saccharomyces* and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including 3-phosphoglycerate kinase or alcohol oxidase, and an origin of replication, termination sequences and the like as desired.

25 A protein of the present invention, once expressed, can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysate. The monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassay or other standard immunoassay techniques.

The sequences encoding proteins of the present invention can also be ligated to various
30 expression vectors for use in transfecting cell cultures of, for instance, mammalian, insect, or plant origin. Illustrative cell cultures useful for the production of the peptides are mammalian cells. Mammalian cell systems often will be in the form of monolayers of cells although

mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been developed in the art, and include the HEK293, BHK21, and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter (*e.g.* the CMV promoter, a HSV *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen *et al.* (1986) *Immunol. Rev.* 89:49), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (*e.g.*, an SV40 large T Ag poly A addition site), and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present invention are available, for instance, from the American Type Culture Collection.

Appropriate vectors for expressing proteins of the present invention in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth and *Drosophila* cell lines such as a Schneider cell line (See, Schneider (1987) *J. Embryol. Exp. Morphol.* 27:353-365).

As with yeast, when higher animal or plant host cells are employed, polyadenylation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague, *et al.* (1983) *J. Virol.* 45:773-781). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo, M., D.M. Glover, ed. (1985) "Bovine Papilloma Virus DNA a Eukaryotic Cloning Vector" in *DNA Cloning Vol. II a Practical Approach*, pp. 213-238 (IRL Press, Arlington, Virginia).

Transfection/Transformation of Cells

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied. Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription and/or translation of the sequence to effect phenotypic changes in the organism. Thus, any method, which provides for effective transformation/transfection may be employed.

A. Plant Transformation

The genes of the present invention can be used to transform any plant. In this manner, genetically modified plants, plant cells, plant tissue, seed, and the like can be obtained.

- 5 Transformation protocols may vary depending on the type of plant cell, i.e. monocot or dicot, targeted for transformation. Suitable methods of transforming plant cells include microinjection (Crossway *et al.* (1986) *BioTechniques* 4:320-334), electroporation (Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium* mediated transformation (Hinchee *et al.* (1988) *Biotechnology* 6:915-921; U.S. Patent No. 5,981,840 (maize); U.S. Patent No. 5,932,782 (sunflower), European Patent No. 0486233 (sunflower); PCT application number WO 98/49332 (sorghum)), direct gene transfer (Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, Sanford *et al.*, U.S. Patent 4,945,050; Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment" In Gamborg and Phillips, eds., *Plant Cell, Tissue and Organ*
- 10 *Culture: Fundamental Methods* (Springer-Verlag, Berlin); McCabe *et al.* (1988) *Biotechnology* 6:923-926); U.S. Patent Number 5,990,387 (maize), U.S. Patent Number 5,886,244 (maize); U.S. Patent Number 5,322,783 (sorghum)). Also see, Weissinger *et al.* (1988) *Annual Rev. Genet.* 22:421-477; Sanford *et al.* (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.* (1988) *Plant Physiol.* 87:671-674 (soybean);
- 20 McCabe *et al.* (1988) *Bio/Technology* 6:923-926 (soybean); Datta *et al.* (1990) *Biotechnology* 8:736-740 (rice); Klein *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein *et al.* (1988) *Biotechnology* 6:559-563 (maize); Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment" in Gamborg and Phillips (eds.) *Plant Cell, Tissue and Organ Culture: Fundamental Methods* (Springer-Verlag, Berlin)
- 25 (maize); Klein *et al.* (1988) *Plant Physiol.* 91:440-444 (maize) Fromm *et al.* (1990) *Biotechnology* 8:833-839 (maize); Hooydaas-Van Slogteren & Hooykaas (1984) *Nature (London)* 311:763-764; Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet *et al.* (1985) In *The Experimental Manipulation of Ovule Tissues* ed. G.P. Chapman *et al.*, pp. 197-209 (Longman, NY) (pollen); Kaeppler *et al.* (1990) *Plant Cell*
- 30 *Reports* 9:415-418; and Kaeppler *et al.* (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505 (electroporation); LI *et al.* (1993) *Plant Cell Reports* 12:250-255, Christou and Ford (1995) *Annals of Botany*

75:745-750 (maize via *Agrobacterium tumefaciens*), and Lec1 transformation (WO 00/28058); all of which are herein incorporated by reference.

The cells, which have been transformed, may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports*, 5:81-84.

5 These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the subject phenotypic characteristics is stable maintained and inherited and then seeds harvested to ensure the desired phenotype or other property has been achieved. One of skill will recognize that after
10 the recombinant expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of number of standard breeding techniques can be used, depending upon the species to be crossed.

In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants.

15 Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plans that would produce the selected phenotype.

20 Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

25 A preferred embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of
30 the seed produced and analyzing the resulting plants produced for altered expression of a polynucleotide of the present invention relative to a control plant (i.e., native, non-transgenic).

Backcrossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

The present invention may be used for transformation of any plant species, including, but not limited to, monocots and dicots. Examples of plant species of interest include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers.

Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Preferably, plants of the present invention are crop plants (for

example, corn, alfalfa, sunflower, *Brassica*, soybean, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.), more preferably corn and soybean plants, yet more preferably corn plants.

B. Transfection of Prokaryotes, Lower Eukaryotes, and Animal Cells

5 Animal and lower eukaryotic (e.g., yeast) host cells are competent or rendered competent for transfection by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate precipitation, fusion of the recipient cells with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextrin, electroporation, biolistics, and
10 micro-injection of the DNA directly into the cells. The transfected cells are cultured by means well known in the art. Kuchler (1997) *Biochemical Methods in Cell Culture and Virology* (Dowden, Hutchinson and Ross, Inc.)

Modulating Polypeptide Levels and/or Composition

15 The present invention further provides a method for modulating (i.e., increasing or decreasing) the concentration or composition of the polypeptides of the present invention in a plant or part thereof. Increasing or decreasing the concentration and/or the composition (i.e., the ratio of the polypeptides of the present invention) in a plant can effect modulation. The method comprised introducing into a plant cell, a recombinant expression cassette comprising
20 a polynucleotide of the present invention as described above to obtain a transformed plant cell, culturing the transformed plant cell under plant cell growing conditions, and inducing or repressing expression of a polynucleotide of the present invention in the plant for a time sufficient to modulate concentration and/or composition in the plant or plant part.

 In some embodiments, the content and/or composition of polypeptides of the present
25 invention in a plant may be modulated by altering, *in vivo* or *in vitro*, the promoter of a gene to up- or down- regulate gene expression. In some embodiments, the coding regions of native genes of the present invention can be altered via substitution, addition, insertion, or deletion to decrease activity of the encoded enzyme. See, e.g., Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868. And in some embodiments, an isolated nucleic acid (e.g., a vector)
30 comprising a promoter sequence is transfected into a plant cell. Subsequently, a plant cell comprising the promoter operably linked to a polynucleotide of the present invention is selected for by means known to those of skill in the art such as, but not limited to, Southern

blot, DNA sequencing, or PCR analysis using primers specific to the promoter and to the gene and detecting amplicons produced therefrom. A plant or plant part altered or modified by the foregoing embodiments is grown under plant forming conditions for a time sufficient to modulate the concentration and/or composition of polypeptides of the present invention in the plant. Plant forming conditions are well known in the art and discussed briefly, *supra*.

In general, concentration or composition is increased or decreased by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned recombinant expression cassette. Modulation in the present invention may occur during and/or subsequent to growth of the plant to the desired stage of development. Modulating nucleic acid expression temporally and/or in particular tissues can be controlled by employing the appropriate promoter operably linked to a polynucleotide of the present invention in, for example, sense or antisense orientation as discussed in greater detail, *supra*. Induction of expression of a polynucleotide of the present invention can also be controlled by exogenous administration of an effective amount of inducing compound. Inducible promoters and inducing compounds, which activate expression from these promoters, are well known in the art. In preferred embodiments, the polypeptides of the present invention are modulated in monocots, particularly maize.

Molecular Markers

The present invention provides a method of genotyping a plant comprising a polynucleotide of the present invention. Optionally, the plant is a monocot, such as maize or sorghum. Genotyping provides a means of distinguishing homologs of a chromosome pair and can be used to differentiate segregants in a plant population. Molecular marker methods can be used for phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., Clark, ed. (1997) *Plant Molecular Biology: A Laboratory Manual*, Chapter 7 (Springer-Verlag, Berlin). For molecular marker methods, see generally, The DNA Revolution by Andrew H. Paterson 1996 (Chapter 2) in: *Genome Mapping in plants* (ed. Andrew H. Paterson) by Academic Press/R.G. Lands Company, Austin, Texas, pp. 7-21.

The particular method of genotyping in the present invention may employ any number of molecular marker analytic techniques such as, but not limited to, restriction fragment length

polymorphism's (RFLPs). RFLPs are the product of allelic differences between DNA restriction fragments resulting from nucleotide sequence variability. As is well known to those of skill in the art, RFLPs are typically detected by extraction of genomic DNA and digestion with a restriction enzyme. Generally, the resulting fragments are separated
5 according to size and hybridized with a probe; single copy probes are preferred. Restriction fragments from homologous chromosomes are revealed. Differences in fragment size among alleles represent an RFLP. Thus, the present invention further provides a means to follow segregation of a gene or nucleic acid of the present invention as well as chromosomal sequences genetically linked to these genes or nucleic acids using such techniques as RFLP
10 analysis. Linked chromosomal sequences are within 50 centiMorgans (cM), often within 40 or 30 cM, preferably within 20 or 10 cM, more preferably within 5, 3, 2, or 1 cM of a gene of the present invention.

In the present invention, the nucleic acid probes employed for molecular marker mapping of plant nuclear genomes selectively hybridize, under selective hybridization
15 conditions, to a gene encoding a polynucleotide of the present invention. In preferred embodiments, the probes are selected from polynucleotides of the present invention. Typically, these probes are cDNA probes or restriction enzyme treated (e.g., *PST I*) genomic clones. The length of the probes is discussed in greater detail, *supra*, but is typically at least 15 bases in length, more preferably at least 20, 25, 30, 35, 40, or 50 bases in length.
20 Generally, however, the probes are less than about 1 kilobase in length. Preferably, the probes are single copy probes that hybridize to a unique locus in haploid chromosome complement. Some exemplary restriction enzymes employed in RFLP mapping are *EcoRI*, *EcoRV*, and *SstI*. As used herein the term "restriction enzyme" includes reference to a composition that recognizes and, alone or in conjunction with another composition, cleaves at a specific
25 nucleotide sequence.

The method of detecting an RFLP comprises the steps of (a) digesting genomic DNA of a plant with a restriction enzyme; (b) hybridizing a nucleic acid probe, under selective hybridization conditions, to a sequence of a polynucleotide of the present of said genomic DNA; (c) detecting therefrom a RFLP. Other methods of differentiating polymorphic (allelic)
30 variants of polynucleotides of the present invention can be had by utilizing molecular marker techniques well known to those of skill in the art including such techniques as: 1) single stranded conformation analysis (SSCA); 2) denaturing gradient gel electrophoresis (DGGE); 3)

RNase protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and 6) allele-specific PCR. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE);
5 heteroduplex analysis (HA); and chemical mismatch cleavage (CMC). Thus, the present invention further provides a method of genotyping comprising the steps of contacting, under stringent hybridization conditions, a sample suspected of comprising a polynucleotide of the present invention with a nucleic acid probe. Generally, the sample is a plant sample, preferably, a sample suspected of comprising a maize polynucleotide of the present invention
10 (e.g., gene, mRNA). The nucleic acid probe selectively hybridizes, under stringent conditions, to a subsequence of a polynucleotide of the present invention comprising a polymorphic marker. Selective hybridization of the nucleic acid probe to the polymorphic marker nucleic acid sequence yields a hybridization complex. Detection of the hybridization complex indicates the presence of that polymorphic marker in the sample. In preferred embodiments,
15 the nucleic acid probe comprises a polynucleotide of the present invention.

UTRs and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA. Positive sequence
20 motifs include translational initiation consensus sequences (Kozak (1987) *Nucleic Acids Res* 15:8125) and the 7-methylguanosine cap structure (Drummond *et al.* (1985) *Nucleic Acids Res.* 13:7375). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing *et al.* (1987) *Cell* 48:691) and AUG sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra*, Rao *et al.* (1988) *Mol. and*
25 *Cell. Biol.* 8:284). Accordingly, the present invention provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present invention can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host
30 such as to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software packages such as "Codon

Preference" available from the University of Wisconsin Genetics Computer Group (see Devereaux *et al.* (1984) *Nucleic Acids Res.* 12:387-395) or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.). Thus, the present invention provides a codon usage frequency characteristic of the coding region of at least one of the polynucleotides of the present invention. The number of polynucleotides that can be used to determine a codon usage frequency can be any integer from 1 to the number of polynucleotides of the present invention as provided herein. Optionally, the polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50, or 100.

10 Sequence Shuffling

The present invention provides methods for sequence shuffling using polynucleotides of the present invention, and compositions resulting therefrom. Sequence shuffling is described in PCT Publication No. WO 96/19256. See also, Zhang, *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509. Generally, sequence shuffling provides a means for generating libraries of polynucleotides having a desired characteristic, which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides, which comprise sequence regions, which have substantial identity and can be homologously recombined *in vitro* or *in vivo*. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be a decreased K_m and/or increased K_{cat} over the wild-type protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140%, or at least 150% of the wild-type value.

Generic and Consensus Sequences

Polynucleotides and polypeptides of the present invention further include those having: (a) a generic sequence of at least two homologous polynucleotides or polypeptides, respectively, of the present invention; and, (b) a consensus sequence of at least three
5 homologous polynucleotides or polypeptides, respectively, of the present invention. The generic sequence of the present invention comprises each species of polypeptide or polynucleotide embraced by the generic polypeptide or polynucleotide, sequence, respectively. The individual species encompassed by a polynucleotide having an amino acid or nucleic acid consensus sequence can be used to generate antibodies or produce nucleic acid
10 probes or primers to screen for homologs in other species, genera, families, orders, classes, phylums, or kingdoms. For example, a polynucleotide having a consensus sequences from a gene family of *Zea mays* can be used to generate antibody or nucleic acid probes or primers to other *Gramineae* species such as wheat, rice, or sorghum. Alternatively, a polynucleotide having a consensus sequence generated from orthologous genes can be used to identify or
15 isolate orthologs of other taxa. Typically, a polynucleotide having a consensus sequence will be at least 9, 10, 15, 20, 25, 30, or 40 amino acids in length, or 20, 30, 40, 50, 100, or 150 nucleotides in length. As those of skill in the art are aware, a conservative amino acid substitution can be used for amino acids, which differ amongst aligned sequence but are from the same conservative amino substitution group as discussed above. Optionally, no more than
20 1 or 2 conservative amino acids are substituted for each 10 amino acid length of consensus sequence.

Similar sequences used for generation of a consensus or generic sequence include any number and combination of allelic variants of the same gene, orthologous, or paralogous sequences as provided herein. Optionally, similar sequences used in generating a consensus
25 or generic sequence are identified using the BLAST algorithm's smallest sum probability (P(N)). Various suppliers of sequence-analysis software are listed in chapter 7 of *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds. Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (Supplement 30). A polynucleotide sequence is considered similar to a reference sequence if the smallest sum
30 probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, or 0.001, and most preferably less than about 0.0001, or 0.00001. Similar polynucleotides can be aligned and a consensus or generic

sequence generated using multiple sequence alignment software available from a number of commercial suppliers such as the Genetics Computer Group's (Madison, WI) PILEUP software, Vector NTI's (North Bethesda, MD) ALIGNX, or Genecode's (Ann Arbor, MI) SEQUENCER. Conveniently, default parameters of such software can be used to generate
5 consensus or generic sequences.

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practices within the scope of the appended claims.

10 EXPERIMENTAL

Example 1. mRNA profiling

The sequences of the present invention were identified as defense inducible by virtue of the induction of their mRNA in the ERE-*avrRxv* callus system which activates the maize pathogen defense system. The mRNA profiling was done using the Affymetrix technology.
15 This technology and the results are described and shown below. The data demonstrates that the sequences of the present invention are induced in the ERE-*avrRxv* system, and that co-induced with them are other known defense related genes. This clearly indicates that their induction is also defense-related.

Materials and Methods:

Construction of ERE-avrRxv Vector:

25 The stable transformation experiments to create estradiol-inducible *avrRxv* expression used a single plasmid construct called "ERE-*avrRxv*". This plasmid contains three tandem plant gene expression units; the estrogen receptor, the estrogen response elements controlling *avrRxv*, and the selectable marker PAT (phosphinothricin acetyltransferase). For the first unit, the nos promoter region (bases 259 to 567 from Bevan *et al.* (1983) *Nucleic Acids Res.*
30 11:369-385) was cloned upstream of the 79 bp tobacco mosaic virus leader omega prime (Gallie *et al.* (1987) *Nucleic Acids Res.* 15:3257-3273) and the first intron of maize alcohol dehydrogenase *ADH1-S* (Dennis *et al.* (1984) *Nucleic Acids Res.* 12:3983-3990). The coding

region for the human estrogen receptor (Tora *et al.* (1989) *EMBO Journal* 8:1981-1986) was inserted between the upstream sequences and the *pinII* terminator. The second unit consists of four pairs of estrogen response element 1/2 sites (EREs) (Klein-Hitpab *et al.* (1986) *Cell* 46: 1053-1061) contained on two copies of the sequence,

- 5 GGCCGCTCGAGTCCAAAGTCAGGTCACAGTGACCTGATCAAAGTTGTCCAAAGTC
AGGTCACAGTGACCTGATCAAAGTTGTCACG (SEQ ID NO:13) (half-sites underlined)
cloned upstream of the minimal *ADH1-S* promoter (bases -89 to +80) and the *ADH1-S* first
intron. The *avrRxv* coding sequence and *pinII* terminator are inserted downstream. The third
unit contains the cauliflower mosaic virus 35S promoter and terminator (bases 6908-7437 and
10 7480-7632 from Franck *et al.* (1980) *Cell* 21: 285-294) controlling expression of a synthetic
coding sequence of phosphinothricin-*N*-acetyltransferase, *pat* (Wohleben *et al.* (1988) *Gene*
70: 25-37) synthesized with plant preferred codons.

Production and Estradiol Treatment of ERE-avrRxv Transgenic Callus and Cell Suspensions:

- 15 For transformation experiments to produce transgenic *ERE-avrRxv* callus, immature
embryos were isolated from greenhouse-grown HiII genotype plants 8-10 days after
pollination. The immature embryos were isolated, cultured and prepared for bombardment as
described above for the transient expression assays. Particle bombardment transformation
was done as described above for immature embryo transformation, except that the
20 transforming DNA was the "ERE-*avrRxv*" construct. One day after 745 embryos were
bombarded, they were transferred to a selection medium similar to the initiation medium but
containing 3 mg/L active ingredient of the herbicide bialaphos® (Meiji Seika Kaisha, LTD,
Yokohama, Japan). From these, 48 transformed colonies were identified between 7 and 9
weeks after bombardment and selected by rapid, healthy growth. Of these 33 were PCR
25 positive for the *avrRxv* gene, among them lines 197 and 186 described herein. Cell
suspensions were generated from ERE-*avrRxv* callus line 197 and control HiII callus by
forcing calli through a 1.5 mm sieve into 250 ml baffled flasks containing 70 ml of liquid
Murashige and Skoog (MS) medium with MS vitamins, 3% sucrose, 2 mg/L 2,4-D. Flasks
were rotated at 140 rpm in the dark at 28°C, and transfers were performed twice weekly, with
30 periodic selections for smaller cell aggregates, with transferred cells kept to approximately 5
ml of packed volume.

Transformed callus and cell suspensions were treated with estradiol to induced *avrRxv* gene expression. For callus treatment the callus tissue was gently broken up into 10-20 mg pieces and then plated on the N6 agar medium described above. Three callus lines were used: HiII::nontransformed control, HiII::ERE-*avrRxv* line 197 and HiII::ERE-*avrRxv* line 186. For the experimentals ethanyl-estradiol (Sigma, St. Louis, MO) was dissolved in 100% ethanol to a 10 mM concentration, and then 34.8 μ l of this stock was dispersed in 4 ml of H₂O for an 87 μ M final concentration. For the controls 34.8 μ l of 100% ethanol was added. The 4 ml of solution was spread over the agar surface of 100 x 25 mm plates, flooding the callus cells. The plates were dried in a sterile flow hood overnight, then covered and further incubated at 23°C in the dark, with reapplication after 72 hours. For cell suspension cultures about 5 ml of cells in a 70 ml of liquid culture received either 70 μ l of 10 mM estradiol in 100% ethanol (final concentration 10 μ M estradiol and 0.1% ethanol) or ethanol only for controls. At the desired timepoints, cells were collected by centrifugation.

mRNA Abundance Profiling using the Affymetrix GeneChip® Technology:

Protocols for preparing *in vitro*-transcribed biotinylated cRNA probes from poly-A⁺ mRNA for Affymetrix GeneChip® gene expression analysis were according to the manufacturer's recommendations (Affymetrix, Santa Clara, CA; Technical Support tel. 1-888-DNA-CHIP), which are described in Wodicka *et al.* (1997) *Nature Biotechnology* 15: 1359-1367. In brief, per sample 2 μ g of poly-A⁺ mRNA, described above in mRNA isolations, was used for the first strand cDNA synthesis. This involved a T7-(dT)₂₄ oligonucleotide primer and reverse transcriptase SuperScript II (Gibco-BRL, Gaithersburg, MD). The second strand synthesis involved *E. coli* DNA Polymerase I (Gibco-BRL, Gaithersburg, MD). The double-stranded cDNA was then cleaned up using phenol/chloroform extraction and phase lock gels (5 Prime-3 Prime, Inc., Boulder, CO) followed by ethanol precipitation. For the *in vitro* transcription to produce cRNA, biotin-11-CTP and biotin-16-UTP, in addition to all four NTPs, were used with T7 transcriptase (Ambion, Austin, TX). The IVT product was cleaned up using Rneasy affinity resin columns (Qiagen, Chatsworth, CA). Labeled *in vitro* transcript (IVT) yields ranged from 62-80 μ g per sample. They were stored at -80°C until use. The IVT products were fragmented in acetate buffer (pH 8.1) at 94°C for 35 minutes prior to chip hybridization.

The GeneChip[®] used in these experiments was constructed by Affymetrix using a set of 1500 maize cDNA EST sequences. In brief, the 1.28 cm x 1.28 cm GeneChip[®] contain a high-density array of 20-mer oligonucleotides affixed to a silicon wafer. These oligonucleotides were synthesized in situ on the silicon wafer by a light-dependent combinatorial chemical synthesis (Fodor *et al.* (1991) *Science* 251: 767-773; Pease *et al.* (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:5022-5026). The oligonucleotide sequences are complementary to the sense strand of Pioneer Hi-Bred's cDNA EST sequences. For each gene there are up to forty 20-mer oligonucleotides synthesized. Twenty of these oligonucleotides are exact matches to different, though sometimes overlapping, regions of the EST sequence.

10 The other 20 oligonucleotides contain one base mismatch in the center, which changes hybridization efficiency. (For a minority of genes there were less than 20 oligo probe pairs, but never less than 15 pairs per gene). The perfect match (PM) and mismatch (MM) oligo probe pairs for each gene are tiled in adjacent regions of the GeneChip. Comparison of the hybridization intensities between different PM oligonucleotides for a given gene, and between

15 PM to MM hybridization intensities for an oligonucleotide pair, are used to determine the overall hybridization to the gene, and hence its level of mRNA abundance in the samples (see Wodicka *et al.* (1997) *Nature Biotechnology* 15: 1359-1367).

Probes of *in vitro* labeled transcript were prepared essentially as described (Wodicka *et al.* (1997) *Nature Biotechnology* 15: 1359-1367) for each of the following four samples: 1)

20 HiII callus control (not estradiol treated); 2) HiII callus estradiol treated; 3) ERE-*avrRxv* callus (line 197) control; and 4) ERE-*avrRxv* callus (line 197) estradiol treated. Twelve μ g IVT for each sample were used per chip hybridization. Each sample was hybridized twice (reps A and B), each rep using a different chip. Hybridization and image scanning conditions, and quantitative analysis and intensity calculations, were essentially as described (Wodicka *et al.* (1997) *Nature Biotechnology* 15: 1359-1367). Comparisons of mRNA abundances were

25 made between each rep of each sample; a total of 8 comparisons. Positive gene expression changes were defined as those showing a 2-fold or more change in at least three of these four comparisons made between the HiII control and ERE-*avrRxv* genotypes. The average and standard error for expression fold changes were calculated from the values of these three or

30 four comparisons.

Results

A high density Affymetrix GeneChip[®] array of some 1500 maize gene sequences was used for surveying mRNA expression changes caused by *avrRxv* expression in transgenic ERE-*avrRxv* callus. It was observed that estradiol treatment of ERE-*avrRxv* callus caused a two-fold or higher change in the expression of 17 genes represented on this array, that were not induced 2-fold or more by estradiol treatment of HiII control callus. The increased expression of six (6) of these sequences is described in Table I. The change in mRNA levels ranged from 2.1 to 33.2 fold.

The extensin-like sequence (SEQ ID NO:1), the cytosolic ascorbate peroxidase-like sequence (SEQ ID NO:5), the metallothionin-like sequence (SEQ ID NO:3), the peroxidase-like sequence (SEQ ID NO:11), the non-specific lipid transfer protein-like sequence (SEQ ID NO:7), and the proteinase inhibitor-like sequence (SEQ ID NO:9) all showed elevated mRNA expression levels in the ERE-*avrRxv* system (see table I).

All of the sequences of the present invention are probable plant defense-related genes, and so these mRNA profiling results further support that a defense reaction is caused by *avrRxv*.

Table I. Gene expression induction in transgenic ERE-*avrRxv* callus treated with estradiol

Gene Name or Description	SEQ ID NO	Fold Change ¹	
		Ave	SE
Non-specific Lipid Transfer Protein-like	SEQ ID NO:7	9.7 ²	1.0
Metallothionein-like	SEQ ID NO:3	4.0	0.9
Extensin-like protein	SEQ ID NO:1	2.6	0.4
Ascorbate Peroxidase-like	SEQ ID NO:5	2.1	0.2
Proteinase Inhibitor-like	SEQ ID NO:9	2.1	0.0
Peroxidase-like	SEQ ID NO:11	12.7	1.1

¹Fold change in mRNA abundance as measured by relative fluorescent intensity of hybridizing signal. Average and SE are calculated from 3 or 4 chip comparisons.

²Average increase is 22.4-fold (SE=1.0), but controls had 2.3-fold (SE=0.4) increase, so relative increase is 9.7-fold.

Example 2. Identification of the gene from a computer homology search

Gene identities can be determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., *et al.*, (1993) *J. Mol. Biol.* 215:403 - 410; see also www.ncbi.nlm.nih.gov/BLAST/) searches under default parameters for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences are analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm. The DNA sequences are translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nature Genetics* 3:266-272) provided by the NCBI. In some cases, the sequencing data from two or more clones containing overlapping segments of DNA are used to construct contiguous DNA sequences.

Sequence alignments and percent identity calculations can be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences can be performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method are KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

The extensin-like sequence (SEQ ID NO:1 and 2) shares about 52% sequence identity at the amino acid level between amino acids 1 to amino acid 81 and about 41% sequence identity between amino acids 20 to 83 to an extensin-like protein from *Arabidopsis* (Acc No. ALO49608).

The metallothionein-like sequence (SEQ ID NO:3 and 4) shares sequence identity to both the metallothionein 2 PFAM family (PF01439) and to the plant PEC family metallothionein PFAM family (PF02068). Specifically, amino acids 1-79 of SEQ ID NO:3 share sequence identity to the metallothionein PFAM domain, while amino acids 2-79 of SEQ ID NO:3 share sequence identity to the plant PEC family metallothionein.

The cytosolic Ascorbate peroxidase-like sequence (SEQ ID NO:5 and 6) shares sequence identity to the Peroxidase PFAM family (PF00141) between about amino acids 19 to 227.

The non-specific lipid transferase (SEQ ID NO:7 and 8) shares about 45% sequence identity from about amino acids 40 to 114 to the dir-1 lipid transferase protein from *Arabidopsis* (Accession No. W73871) and about 46% sequence identity from about amino acids 82 to 128 to the non-specific lipid transfer-like protein from *Phaseolus vulgaris* (Accession No. AAC49370). The sequence further share sequence identity to the protease inhibitor/seed storage family of PFAM (tryp_alpha_amyl) (PF00234) from about amino acid 38 to about amino acid 82.

The proteinase inhibitor-like sequence (SEQ ID NO:9 and 10) shares sequence identity to the Bowman-Birk serine protease inhibitor PFAM family (PF00228) from about amino acids 50 to 104.

The peroxidase-like sequence (SEQ ID NO:11 and 12) shares sequence identity to the peroxidase PFAM family 1 (PF00141) from about amino acid 39 to about amino acid 296.

Example 3. Transformation and Regeneration of Transgenic Plants

Immature maize embryos from greenhouse donor plants are bombarded with a plasmid containing the defense-induced sequences of the present invention operably linked to a ubiquitin promoter and the selectable marker gene PAT (Wohlleben *et al.* (1988) *Gene* 70:25-37), which confers resistance to the herbicide Bialaphos. Alternatively, the selectable marker gene is provided on a separate plasmid. Transformation is performed as follows. Media recipes follow below.

Preparation of Target Tissue

The ears are husked and surface sterilized in 30% Clorox bleach plus 0.5% Micro detergent for 20 minutes, and rinsed two times with sterile water. The immature embryos are excised and placed embryo axis side down (scutellum side up), 25 embryos per plate, on 560Y medium for 4 hours and then aligned within the 2.5-cm target zone in preparation for bombardment.

Preparation of DNA

This plasmid DNA plus plasmid DNA containing a PAT selectable marker is precipitated onto 1.1 μm (average diameter) tungsten pellets using a CaCl_2 precipitation procedure as follows:

100 µl prepared tungsten particles in water
10 µl (1 µg) DNA in Tris EDTA buffer (1 µg total DNA)
100 µl 2.5 M CaCl₂
10 µl 0.1 M spermidine

5

Each reagent is added sequentially to the tungsten particle suspension, while maintained on the multitube vortexer. The final mixture is sonicated briefly and allowed to incubate under constant vortexing for 10 minutes. After the precipitation period, the tubes are centrifuged briefly, liquid removed, washed with 500 ml 100% ethanol, and centrifuged for 30 seconds. Again the liquid is removed, and 105 µl 100% ethanol is added to the final tungsten particle pellet. For particle gun bombardment, the tungsten/DNA particles are briefly sonicated and 10 µl spotted onto the center of each macrocarrier and allowed to dry about 2 minutes before bombardment.

15 Particle Gun Treatment

The sample plates are bombarded at level #4 in particle gun #HE34-1 or #HE34-2. All samples receive a single shot at 650 PSI, with a total of ten aliquots taken from each tube of prepared particles/DNA.

20 Subsequent Treatment

Following bombardment, the embryos are kept on 560Y medium for 2 days, then transferred to 560R selection medium containing 3 mg/liter Bialaphos, and subcultured every 2 weeks. After approximately 10 weeks of selection, selection-resistant callus clones are transferred to 288J medium to initiate plant regeneration. Following somatic embryo maturation (2-4 weeks), well-developed somatic embryos are transferred to medium for germination and transferred to the lighted culture room. Approximately 7-10 days later, developing plantlets are transferred to 272V hormone-free medium in tubes for 7-10 days until plantlets are well established. Plants are then transferred to inserts in flats (equivalent to 2.5" pot) containing potting soil and grown for 1 week in a growth chamber, subsequently grown an additional 1-2 weeks in the greenhouse, then transferred to classic 600 pots (1.6 gallon) and grown to maturity. Plants are monitored and scored for and altered level of expression of the

30

defense-inducible sequence of the invention. Alternatively, plants can be assayed for a modulation in disease resistance, or a modulation in extensin-like activity, peroxidase-like activity, a metallothionein-like activity, or a peroxidase-like activity.

5

Bombardment and Culture Media

Bombardment medium (560Y) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 120.0 g/l sucrose,
10 1.0 mg/l 2,4-D, and 2.88 g/l L-proline (brought to volume with D-I H₂O following adjustment to pH 5.8 with KOH); 2.0 g/l Gelrite (added after bringing to volume with D-I H₂O); and 8.5 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature).

Selection medium (560R) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 30.0 g/l sucrose, and
15 2.0 mg/l 2,4-D (brought to volume with D-I H₂O following adjustment to pH 5.8 with KOH); 3.0 g/l Gelrite (added after bringing to volume with D-I H₂O); and 0.85 mg/l silver nitrate and 3.0 mg/l bialaphos(both added after sterilizing the medium and cooling to room temperature).

Plant regeneration medium (288J) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l
20 pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H₂O) (Murashige and Skoog (1962) *Physiol. Plant.* 15:473), 100 mg/l myo-inositol, 0.5 mg/l zeatin, 60 g/l sucrose, and 1.0 ml/l of 0.1 mM abscisic acid (brought to volume with polished D-I H₂O after adjusting to pH 5.6); 3.0 g/l Gelrite (added after bringing to volume with D-I H₂O); and 1.0 mg/l indoleacetic acid and 3.0 mg/l bialaphos (added after sterilizing the medium and cooling
25 to 60°C). Hormone-free medium (272V) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g/l nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H₂O), 0.1 g/l myo-inositol, and 40.0 g/l sucrose (brought to volume with polished D-I H₂O after adjusting pH to 5.6); and 6 g/l bacto-agar (added after bringing to volume with polished D-I H₂O), sterilized
30 and cooled to 60° C.

Example 4. *Agrobacterium*-mediated Transformation

For *Agrobacterium*-mediated transformation of maize with a defense induced preferably the method of Zhao is employed (U.S. Patent No. 5,981,840, and PCT patent publication WO98/32326; the contents of which are hereby incorporated by reference). Briefly, immature embryos are isolated from maize and the embryos contacted with a suspension of *Agrobacterium*, where the bacteria are capable of transferring the defense-inducible nucleotide sequences to at least one cell of at least one of the immature embryos (step 1: the infection step). In this step the immature embryos are preferably immersed in an *Agrobacterium* suspension for the initiation of inoculation. The embryos are co-cultured for a time with the *Agrobacterium* (step 2: the co-cultivation step). Preferably the immature embryos are cultured on solid medium following the infection step. Following this co-cultivation period an optional "resting" step is contemplated. In this resting step, the embryos are incubated in the presence of at least one antibiotic known to inhibit the growth of *Agrobacterium* without the addition of a selective agent for plant transformants (step 3: resting step). Preferably the immature embryos are cultured on solid medium with antibiotic, but without a selecting agent, for elimination of *Agrobacterium* and for a resting phase for the infected cells. Next, inoculated embryos are cultured on medium containing a selective agent and growing transformed callus is recovered (step 4: the selection step). Preferably, the immature embryos are cultured on solid medium with a selective agent resulting in the selective growth of transformed cells. The callus is then regenerated into plants (step 5: the regeneration step), and preferably calli grown on selective medium are cultured on solid medium to regenerate the plants.

Example 5. Soybean Embryo Transformation

Soybean embryos are bombarded with a plasmid containing the defense-inducible sequence operably linked to the Scp1 promoter (U.S. Patent No. 6,072,050) as follows. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface-sterilized, immature seeds of the soybean cultivar A2872, are cultured in the light or dark at 26°C on an appropriate agar medium for six to ten weeks. Somatic embryos producing secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos that multiplied as early, globular-staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 ml liquid media on a rotary shaker, 150 rpm, at 26° C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 ml of liquid medium.

5 Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein *et al.* (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic PDS1000/HE instrument (helium retrofit) can be used for these transformations.

10 A selectable marker gene that can be used to facilitate soybean transformation is a transgene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell *et al.* (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz *et al.* (1983) *Gene* 25:179-188), and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The expression cassette comprising the defense-inducible sequence operably linked to the Scp1 promoter can be
15 isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µl of a 60 mg/ml 1 µm gold particle suspension is added (in order): 5 µl DNA (1 µg/µl), 20 µl spermidine (0.1 M), and 50 µl CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed.
20 The DNA-coated particles are then washed once in 400 µl 70% ethanol and resuspended in 40 µl of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five microliters of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an
25 empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi, and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be
30 divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post-bombardment with fresh media containing 50 mg/ml hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post-bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

Example 6. Sunflower Meristem Tissue Transformation

Sunflower meristem tissues are transformed with an expression cassette containing the defense-induced sequence operably linked to the Scp1 promoter as follows (see also European Patent Number EP 0 486233, herein incorporated by reference, and Malone-Schoneberg *et al.* (1994) *Plant Science* 103:199-207). Mature sunflower seed (*Helianthus annuus* L.) are dehulled using a single wheat-head thresher. Seeds are surface sterilized for 30 minutes in a 20% Clorox bleach solution with the addition of two drops of Tween 20 per 50 ml of solution. The seeds are rinsed twice with sterile distilled water.

Split embryonic axis explants are prepared by a modification of procedures described by Schrammeijer *et al.* (Schrammeijer *et al.* (1990) *Plant Cell Rep.* 9: 55-60). Seeds are imbibed in distilled water for 60 minutes following the surface sterilization procedure. The cotyledons of each seed are then broken off, producing a clean fracture at the plane of the embryonic axis. Following excision of the root tip, the explants are bisected longitudinally between the primordial leaves. The two halves are placed, cut surface up, on GBA medium consisting of Murashige and Skoog mineral elements (Murashige *et al.* (1962) *Physiol. Plant.*, 15: 473-497), Shepard's vitamin additions (Shepard (1980) in *Emergent Techniques for the Genetic Improvement of Crops* (University of Minnesota Press, St. Paul, Minnesota), 40 mg/l adenine sulfate, 30 g/l sucrose, 0.5 mg/l 6-benzyl-aminopurine (BAP), 0.25 mg/l indole-3-acetic acid (IAA), 0.1 mg/l gibberellic acid (GA₃), pH 5.6, and 8 g/l Phytagar.

The explants are subjected to microprojectile bombardment prior to *Agrobacterium* treatment (Bidney *et al.* (1992) *Plant Mol. Biol.* 18: 301-313). Thirty to forty explants are

placed in a circle at the center of a 60 X 20 mm plate for this treatment. Approximately 4.7 mg of 1.8 mm tungsten microprojectiles are resuspended in 25 ml of sterile TE buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0) and 1.5 ml aliquots are used per bombardment. Each plate is bombarded twice through a 150 mm nytex screen placed 2 cm above the samples in a PDS

5 1000® particle acceleration device.

Disarmed *Agrobacterium tumefaciens* strain EHA105 is used in all transformation experiments. A binary plasmid vector comprising the expression cassette described above is introduced into *Agrobacterium* strain EHA105 via freeze-thawing as described by Holsters *et al.* (1978) *Mol. Gen. Genet.* 163:181-187. This plasmid further comprises a kanamycin
10 selectable marker gene (i.e., *nptII*). Bacteria for plant transformation experiments are grown overnight (28°C and 100 RPM continuous agitation) in liquid YEP medium (10 gm/l yeast extract, 10 gm/l Bactopeptone, and 5 gm/l NaCl, pH 7.0) with the appropriate antibiotics required for bacterial strain and binary plasmid maintenance. The suspension is used when it reaches an OD₆₀₀ of about 0.4 to 0.8. The *Agrobacterium* cells are pelleted and resuspended
15 at a final OD₆₀₀ of 0.5 in an inoculation medium comprised of 12.5 mM MES pH 5.7, 1 gm/l NH₄Cl, and 0.3 gm/l MgSO₄.

Freshly bombarded explants are placed in an *Agrobacterium* suspension, mixed, and left undisturbed for 30 minutes. The explants are then transferred to GBA medium and co-cultivated, cut surface down, at 26°C and 18-hour days. After three days of co-cultivation, the
20 explants are transferred to 374B (GBA medium lacking growth regulators and a reduced sucrose level of 1%) supplemented with 250 mg/l cefotaxime and 50 mg/l kanamycin sulfate. The explants are cultured for two to five weeks on selection and then transferred to fresh 374B medium lacking kanamycin for one to two weeks of continued development. Explants with differentiating, antibiotic-resistant areas of growth that have not produced shoots suitable for
25 excision are transferred to GBA medium containing 250 mg/l cefotaxime for a second 3-day phytohormone treatment. Leaf samples from green, kanamycin-resistant shoots are assayed for the presence of NPTII by ELISA and for the presence of transgene expression by assaying for the activity of the defense inducible sequences.

NPTII-positive shoots are grafted to Pioneer® hybrid 6440 *in vitro*-grown sunflower
30 seedling rootstock. Surface sterilized seeds are germinated in 48-0 medium (half-strength Murashige and Skoog salts, 0.5% sucrose, 0.3% gelrite, pH 5.6) and grown under conditions

described for explant culture. The upper portion of the seedling is removed, a 1 cm vertical slice is made in the hypocotyl, and the transformed shoot inserted into the cut. The entire area is wrapped with parafilm to secure the shoot. Grafted plants can be transferred to soil following one week of *in vitro* culture. Grafts in soil are maintained under high humidity conditions followed by a slow acclimatization to the greenhouse environment. Transformed sectors of T₀ plants (parental generation) maturing in the greenhouse are identified by NPTII ELISA and/or by the analysis of the activity of the defense induced sequences in the leaf extracts while transgenic seeds harvested from NPTII-positive T₀ plants are identified by the analysis of the activity the defense induced sequences in small portions of dry seed cotyledon.

10 An alternative sunflower transformation protocol allows the recovery of transgenic progeny without the use of chemical selection pressure. Seeds are dehulled and surface-sterilized for 20 minutes in a 20% Clorox bleach solution with the addition of two to three drops of Tween 20 per 100 ml of solution, then rinsed three times with distilled water. Sterilized seeds are imbibed in the dark at 26°C for 20 hours on filter paper moistened with
15 water. The cotyledons and root radical are removed, and the meristem explants are cultured on 374E (GBA medium consisting of MS salts, Shepard vitamins, 40 mg/l adenine sulfate, 3% sucrose, 0.5 mg/l 6-BAP, 0.25 mg/l IAA, 0.1 mg/l GA, and 0.8% Phytagar at pH 5.6) for 24 hours under the dark. The primary leaves are removed to expose the apical meristem, around 40 explants are placed with the apical dome facing upward in a 2 cm circle in the
20 center of 374M (GBA medium with 1.2% Phytagar), and then cultured on the medium for 24 hours in the dark.

 Approximately 18.8 mg of 1.8 µm tungsten particles are resuspended in 150 µl absolute ethanol. After sonication, 8 µl of it is dropped on the center of the surface of macrocarrier. Each plate is bombarded twice with 650 psi rupture discs in the first shelf at 26
25 mm of Hg helium gun vacuum.

 The plasmid of interest is introduced into *Agrobacterium tumefaciens* strain EHA105 via freeze thawing as described previously. The pellet of overnight-grown bacteria at 28 °C in a liquid YEP medium (10 g/l yeast extract, 10 g/l Bactopeptone, and 5 g/l NaCl, pH 7.0) in the presence of 50 µg/l kanamycin is resuspended in an inoculation medium (12.5 mM 2-mM
30 2-(N-morpholino) ethanesulfonic acid, MES, 1 g/l NH₄Cl and 0.3 g/l MgSO₄ at pH 5.7) to reach a final concentration of 4.0 at OD 600. Particle-bombarded explants are transferred to

GBA medium (374E), and a droplet of bacteria suspension is placed directly onto the top of the meristem. The explants are co-cultivated on the medium for 4 days, after which the explants are transferred to 374C medium (GBA with 1% sucrose and no BAP, IAA, GA3 and supplemented with 250 µg/ml cefotaxime). The plantlets are cultured on the medium for about two weeks under 16-hour day and 26 °C incubation conditions.

Explants (around 2 cm long) from two weeks of culture in 374C medium are screened for defense induced activity using assays known in the art. After positive (i.e., for defense-inducible expression) explants are identified, those shoots that fail to exhibit defense-inducible activity are discarded, and every positive explant is subdivided into nodal explants. One nodal explant contains at least one potential node. The nodal segments are cultured on GBA medium for three to four days to promote the formation of auxiliary buds from each node. Then they are transferred to 374C medium and allowed to develop for an additional four weeks. Developing buds are separated and cultured for an additional four weeks on 374C medium. Pooled leaf samples from each newly recovered shoot are screened again by the appropriate protein activity assay. At this time, the positive shoots recovered from a single node will generally have been enriched in the transgenic sector detected in the initial assay prior to nodal culture.

Recovered shoots positive for defense-inducible expression are grafted to Pioneer hybrid 6440 *in vitro*-grown sunflower seedling rootstock. The rootstocks are prepared in the following manner. Seeds are dehulled and surface-sterilized for 20 minutes in a 20% Clorox bleach solution with the addition of two to three drops of Tween 20 per 100 ml of solution, and are rinsed three times with distilled water. The sterilized seeds are germinated on the filter moistened with water for three days, then they are transferred into 48 medium (half-strength MS salt, 0.5% sucrose, 0.3% gelrite pH 5.0) and grown at 26 °C under the dark for three days, then incubated at 16-hour-day culture conditions. The upper portion of selected seedling is removed, a vertical slice is made in each hypocotyl, and a transformed shoot is inserted into a V-cut. The cut area is wrapped with parafilm. After one week of culture on the medium, grafted plants are transferred to soil. In the first two weeks, they are maintained under high humidity conditions to acclimatize to a greenhouse environment.

Example 7: Anti-fungal and Anti-bacterial assays

Anti-fungal Assays: *F. graminearum* is grown in half-strength CM-cellulose-yeast extract broth (7.5 g of CM-cellulose, 0.5 g of yeast extract, 0.25 g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5 g of NH_4NO_3 , and 0.5 g of KH_2PO_4 per liter of distilled water). Cultures are shaken at 200 rpm at 28°C in the light. After 7 days, cultures are filtered through two layers of sterile cheesecloth and the resulting filtrate is passed through a Nalgene 0.45- μm disposable filter unit. Conidia (spores) are collected on the membrane, washed with sterile distilled water, and resuspended in a small volume of sterile water. Conidia are counted with a hemocytometer and stored at 4°C for not longer than 1 month. *A. longipes* cultures are grown on carrot agar at 28°C under continuous fluorescent light, and *F. moniliforme* and *A. flavus* are grown on oatmeal agar at 28°C under ambient light. For these three fungi, conidia are collected by scraping a sterile inoculating loop across the surface of the plate. Concentrated suspensions are made in sterile water with 0.1% Tween 20. Conidia are counted with a hemocytometer and used immediately. For an assay, fungal spore suspensions are diluted to give a concentration of 250 spores/90 μl of dilute culture medium (0.037 g of NaCl, 0.0625 g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.25 g of calcium nitrate, 2.5 g of glucose, 0.25 g of yeast extract, 0.125 g of casein hydrolysate (enzyme), and 0.125 g of casein hydrolysate (acid) in 7.5 mM sodium phosphate buffer, pH 5.8).

For *Sclerotinia* cultures, mycelia are grown on cellophane discs (52 mm) overlain on V8 agar. When hyphal growth reaches the margin of the disc, the cellophane is removed and the mycelium is dislodged by vortexing in 10 ml of diluted culture medium, followed by filtration through two layers of cheesecloth. Hyphal pieces are washed by centrifugation at 2000 rpm for 5 min and are resuspended in diluted growth medium to give a concentration of approximately 50 hyphal pieces/90 μl .

To perform anti-fungal assays, 10 μl of test material in water or 0.01% acetic acid are added to wells of a microtiter plate. Ninety microliters of spores or hyphal pieces are added and mixed. Plates are covered and incubated at 28° in the dark for 24-48 h. Growth is evaluated visually using an inverted microscope, and a scale of 0-4 is used to rate the effect of added peptide (0 = no observable inhibition relative to water control; 1 = slight inhibition; 2 = substantial inhibition; 3 = almost complete inhibition; 4 = complete inhibition).

Anti-bacterial Assays. Cultures are grown to midlog phase (*E. coli* in LB broth and *C. nebraskense* in NBY) and are then harvested by centrifugation (2000 x g for 10 min). Cells

are washed with 10 mM sodium phosphate buffer, pH 5.8 (*C. nebraskense*) or pH 6.5 (*E. coli*) by centrifugation and then colony forming units are estimated spectrophotometrically at 600 nm with previously established colony forming unit-optical density relationships used as a reference.

5 Assays for bactericidal activity are performed by incubating 10^5 bacterial colony forming units in 90 μ l with 10 ml of peptide (or water for control). After 60 min at 37°C (*E. coli*) or 25°C (*C. nebraskense*), four serial, 10-fold dilutions are made in sterile phosphate buffer. Aliquots of 100 μ l are plated on LB or NBY plates, using 1 or 2 plates/dilution. Resulting colonies are counted, and the effect of the peptide is expressed as percent of initial
10 colony count (Selsted *et al.* (1984) *Infect. Immun.* 45:150-154).

 Assays for bacteriostatic activity are performed by incubating 10^5 bacteria with MBP-1 in 200 μ l of dilute medium (1 part NBY broth to 4 parts 10 mM sodium phosphate, pH 5.8) in microtiter plate wells. Plates are covered, sealed, and incubated at 28°C. Growth is monitored spectrophotometrically at 600 nm. After 41 h controls will have grown sufficiently
15 (optical density >0.20) to measure effect of peptide as percent of control.

Example 8: Protease Inhibition Assays

 Apparent K_i values are determined for the wild type proteinase inhibitor-like sequences of the invention using the equation $V_o/V_i = 1 + [I]/K_{i(app)}$, where V_o is the reaction
20 rate in the absence of inhibitor, and V_i is the reaction rate in the presence of inhibitor (Nicklin and Barrett (1984) *Biochem J.* 223:245-249). Reactions without inhibitor are started by addition of substrate, and the linear increase in absorbance at 405 nm is monitored over time and the reaction rate calculated from the slope. A known quantity of inhibitor is then added to the same reaction, and the new reaction rate is determined. The following proteases can be
25 used: bovine pancreatic chymotrypsin, bovine pancreatic trypsin, porcine pancreatic elastase and subtilisin Carlsberg from *Bacillus licheniformis* (all from Sigma). Assays are done at 37°C for chymotrypsin, and at 25°C for the other proteases. Reaction volumes are typically 200 μ l. The following substrates are used at a concentration of 1 mM: *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (Sigma) for chymotrypsin and subtilisin, *N*-benzoyl-2-Ile-Glu-Gly-Arg-*p*-nitroanilide (Chromogenix S-2222) for trypsin and *N*-succinyl-Ala-Ala-Ala-*p*-
30 nitroanilide (Sigma) for elastase. Chymotrypsin, elastase and subtilisin assays are done in 200

mM Tris-HCl, pH 8.0, with 1 μ M bovine serum albumin included. Trypsin assays are done in 50 mM Tris-HCl, 2 mM NaCl, 2 mM CaCl₂, 0.005% TritonX-100, pH 7.5.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are indicative of the level of those skilled in the art to which this invention pertains. All publications, patents, and patent applications are hereby incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

10